

TARTU RIIKLIK ÜLIKOOL



LINGUISTICA

TARTU RIIKLIKU ÜLIKOOLI TOIMETISED  
УЧЕНЫЕ ЗАПИСКИ  
ТАРТУСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА  
ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS  
ALUSTATUD 1893.a. VIHK 453 ВЫПУСК ОСНОВАН В 1893.g.

ÜLD- JA RAKENDUSLINGVISTIKA  
PROBLEEME  
ПРОБЛЕМЫ ОБЩЕЙ И ПРИКЛАДНОЙ  
ЛИНГВИСТИКИ

L I N G U I S T I C A X

TARTU 1978

**Toimetuse kolleegium:**

**Paul Ariste (esimees), Helgi Liin, Oleg Mutt, Rita Tasa,  
Juhan Tuldava (vastutav toimetaja)**

**Редакционная коллегия:**

**Пауль Аристе (председ.), Хельги Лийн, Олег Мутт, Рита Таза,  
Ехан Тулдава (отв. редактор)**

Ученые записки Тартуского государственного университета.  
Выпуск 453. ПРОБЛЕМЫ ОБЩЕЙ И ПРИКЛАДНОЙ ЛИНГВИСТИКИ.  
Лингвистика X. На эстонском, русском, немецком и англий-  
ском языках. Резюме на разных языках. Тартуский государ-  
ственный университет. СССР, г. Тарту, ул. Киликооли, 18.  
Ответственный редактор Ю. Тулдава. Сдано в печать 24.03.  
1978. Бумага печатная 30x45 1/4. Печ. листов 10,0. Учет-  
но-издат. листов 9,63. Тираж 500. МВ 03823. Типография  
ТГУ, СССР, г. Тарту, ул. Пялсони, 14. Зак. № 353.  
Цена 1 руб. 40 коп.

Kogumiku "Linguistica" kümmandas väljaandes ("Üld- ja rakenduslingvistika probleemid") on avaldatud Tartu Riikliku Ülikooli õppejõudude ja külalisautorite artiklid. O. Mutt annab ülevaate viimase aja tähtsamatest uurimistöödest anglistika valdkonnas. B. Gasparov ja V. Paperno vaatlevad oma artiklis inglise kõnekeele tüpoloogiat. U. Lehtsalu ja G. Liiv käsitleva inglise-eesti tõlkeküsimusi aktuaalse liigenduse seisukohast. T.-R. Viitso artikkel on pühendatud fonoloogiliste rekonstruktsioonide ja algkeele vahekorrale. A. Dulichenko käsitleb oma kirjutises lingvistika metakeele arengut. H. Üim analüüsib tegevusverbide semantilist struktuuri. Nn. hinnangukategooriat rootsikeelse materjali põhjal vaatleb I. Mullaama, kuma L. Paavo vaatleb sõnakasutuse ja ideoloogia seoseid Saksa DV-s ja Saksa FV-s. K. Uustalu kirjutises vaadeldakse eestikeelsete kohanime laenamist saksa keelde 17. sajandi ürikute varal. Ülejäänud artiklid kuuluvad keelestatistika ja stilii statistilise uurimise valdkonda. Külalisautorid J. Orlov ja I. Nadareišvili (Tbilisi) esitavad kompleksse keelestatistilise uurimise nõuded ja vaatlevad vastavaid meetodeid. A. Šaikovitš (Moskva) käsitleb poeetilise stroofi positsioonilise analüüsi põhimõtteid ja rakendab neid inglise soneti ja Puškini loomingu uurimisel. V. Bogodist (Uljanovsk) vaatleb oma artiklis prantsuskeelse teksti semantilis-pragmatilise informatsiooni mõtlemist ja seostab seda keeleõpetamise metoodikaga. J. Tuldava artikkel on pühendatud eesti keele ühesilbiliste sõnade struktuuri kvantitatiivsele uurimisele.

В десятом выпуске сборника Linguistica ("Проблемы общей и прикладной лингвистики") опубликованы статьи преподавателей Тартуского государственного университета и исследователей из других городов. В статье О. Мутта дается обзор зарубежных работ по англистике за последнее время. Б. Гаспаров и В. Паперно рассматривают в своей совместной статье некоторые вопросы типологии английской разговорной речи. Статья У. Лехтсалу и Г. Лиив посвящена сопоставлению семантической структуры предложения в английском и эстонском языках. Т.-Р. Вийтсо пишет о неоднозначности фонологических реконструкций. В статье А. Дуличенко рассматривается проблема эволюции метаязыка современной лингвистики. Х. Ыйи анализирует семантическую структуру слов, выражающих действия. Темой работы И. Мullaama является изучение стилистической категории оценочности на материале общественно-политической лексики шведского языка. Л. Пааво пишет об идеологической связанности новой лексики в ГДР и ФРГ. В статье К. Уусталу рассматриваются условия заимствования эстонских топонимических названий в немецкий язык по материалам хроник 17-го века. Остальные статьи относятся к количественной лингвистике и стилистике. Д. Орлов и И. Надарейшвили (Тбилиси) представляют метод полной фиксации текста при лингвостатистическом анализе. А. Шайкевич (Москва) рассматривает принципы позиционного анализа поэтической строфы и применяет этот метод при сопоставительном исследовании елизаветинского сонета и онегинской строфы. В. Богодист (Ульяновск) изучает информативность слова для носителей языка и для иностранцев на материале французского языка. Статья Д. Тулдава посвящена количественному исследованию односложных слов в эстонском языке.

In dem zehnten Heft der Schriftenreihe "Linguistica" ("Probleme der allgemeinen und angewandten Sprachwissenschaft") werden die Arbeiten der Lehrkräfte der Tartuer Staatlichen Universität und anderer höherer Lehranstalten der SU veröffentlicht. In seiner Abhandlung gibt O. Mutt einen Überblick über die neuesten wichtigsten Untersuchungsergebnisse auf dem Gebiet der Anglistik, Gegenstand der Untersuchung von B. Gasparow und V. Paperno ist die Typologie der englischen Umgangssprache, U. Lehtsalu und G. Liiv gehen vom Standpunkt der aktuellen Gliederung einigen Problemen bei der Übersetzung aus dem Englischen ins Estnische nach. Der Artikel von T.-R. Viitso ist dem Verhältnis zwischen den phonologischen Rekonstruktionen und der Ursprache gewidmet. A. Duličenko verfolgt in seinem Beitrag die Entwicklung der linguistischen Metasprache. H. Õim analysiert die semantische Struktur der Handlungsverben. Anhand des schwedischsprachigen Materials betrachtet I. Mullamaa in seiner Arbeit die Probleme der sogenannten Wertungskategorie. L. Paavo legt in ihrem Artikel einige Probleme der Ideologiegelundenheit der neuesten Lexik in der DDR und in der BRD dar. K. Uustalu behandelt anhand einer Chronik des 17. Jh. die Entlehnungen der estnischen Ortsnamen in der deutschen Sprache. Die übrigen Arbeiten gehören zum Bereich der quantitativen Untersuchung der Sprache. Die Gastautoren J. Orlov und I. Nadareishvili (Tbilisi) legen die Forderungen dar, die an die komplexe sprachstatistische Untersuchung gestellt werden und betrachten die entsprechenden Methoden. A. Shaikevitch (Moskau) behandelt die Prinzipien der positionellen Analyse der poetischen Strophe und wendet sie bei der vergleichenden Untersuchung des englischen Sonetts und der Strophe von Puschkin an. V. Bogodist (Uljanowsk) geht in seinem Artikel den Problemen der Messung der semantisch-pragmatischen Information der französischen Texte nach und verbindet sie mit der Methodik des Sprachunterrichts. Die Arbeit von J. Tuldava ist der systematischen Untersuchung der quantitativen Struktur der estnischen einsilbigen Wörter gewidmet.

This tenth issue of "Linguistica" ("Problems of General and Applied Linguistics") consists of papers by authors from Tartu State University and elsewhere. O. Mutt surveys some current research projects in the field of English studies. The typology of spoken English is examined by B. Gasparov and V. Paperno. In a joint article U. Lehtsalu and G. Liiv discuss some problems of translating from English into Estonian from the viewpoint of functional perspective. T.-R. Viitso deals with the relations between phonological reconstruction and parent language. A. Duličenko investigates an aspect of the evolution of the metalanguage of linguistics. H. Õim analyses the semantic structure of verbs of action. The category of evaluation is examined on the basis of Swedish material by I. Mullamaa, while L. Paavo deals with the ties between neologisms and ideology in the GDR and the FRG. K. Uustalu's paper is about the borrowing of Estonian toponyms in German written records of the 17th century. The remaining papers are devoted to quantitative linguistics. Of the guest authors Yu. Orlov and I. Nadareishvili (Tbilisi) lay down the requirements for a complex linguo-statistical study and examine the relevant methods. A. Shaikevitch (Moscow) deals with the principles of the positional analysis of the poetic strophe and applies them to the Elizabethan sonnet and Pushkin's "Eugene Onegin". V. Bogodist (Ulyanovsk) examines the measurement of semantic-pragmatic information in French texts and links this to language teaching methodology. J. Tuldava (Tartu) presents a systematic investigation of the quantitative structure of Estonian monosyllables.

# ИНФОРМАТИВНОСТЬ СЛОВА ДЛЯ НОСИТЕЛЕЙ ЯЗЫКА И ДЛЯ ИНОСТРАНЦЕВ

В.И. Богодист

Всякий осмысленный текст обязательно несет синтаксическую (структурную) и семантическую (содержательную) информацию. В зависимости от характера текста в нем может присутствовать эффективная (эстетическая) информация. Если сообщение является полезным (ценным) для получателя, то это означает, что в нем заключена и прагматическая информация.

В настоящее время разработаны приемы измерения только синтаксической и семантико-прагматической информации, содержащейся в тексте и составляющих его лингвистических единицах.

Целью настоящей работы является описание эксперимента по измерению семантико-прагматической информации, несомой некоторыми французскими словами, и анализ полученных данных в плане возможного их использования в методике преподавания французского языка на специальном отделении педагогического института.

## Эксперимент

Технология эксперимента, впервые описанная Р.Г. Пиотровским (1970) и апробированная на материале ряда языков (см. Богодист, Георгиев, Пестунов, Пиотровский, Райтар, 1975; Bogodist et al., 1975) следующая.

Пусть имеется текст, представляющий собой цепочку слов

$$w_1, w_2, w_3, w_4 \dots, w_k \dots,$$

и нам нужно оценить количество информации, содержащейся в слове  $w_1$ .

Для решения этой задачи необходимо провести коллективное угадывание сегмента текста  $w_2 + w_k^*$ . Первый раз коллективу

\* Коллективное угадывание осуществляется следующим образом. Группе информантов предлагается побуквенно отгадывать сегмент, следующий за контрольным словом. Каждый из испытуемых независимо от других записывает наиболее вероятное с его точки зрения буквенное продолжение для  $n$ -го шага текста, затем экспериментатор сообщает действительную букву и испытуемые переходят к угадыванию следующей буквы.

информантов сообщается слово  $w_1$ , второй раз угадывание начинается прямо со слова  $w_2$ . Полученная в первом случае от контрольного сегмента информация, равная неопределенности угадывания указанного сегмента  $I(w_2 \div w_k / w_1) = H(w_2 \div w_k / w_1)$ , будет меньше информации и неопределенности  $I(w_2 \div w_k) = H(w_2 \div w_k)$ , полученной при условии, что испытуемым было неизвестно первое слово  $w_1$ . Разность  $I(w_2 \div w_k) - I(w_2 \div w_k / w_1)$  представляет собой количественную оценку той информации, которая содержится в слове  $w_1$ .

Именно эта информация уменьшила неопределенность контрольного сегмента и облегчила первое угадывание. Назовем ее семантико-прагматической или смысловой информацией (Исм).

### Математическая обработка результатов эксперимента

Как показывает процедура, Исм измеряется с помощью синтаксической информации (I). Расчет последней производится по второй формуле Шеннона (1963):

$$I = H = - \sum_{n=1}^S p_n \log_2 p_n \quad (1)$$

Известные нам объяснения этой формулы доступны скорее математикам, чем лингвистам и методистам. В связи с этим покажем на примере, как ею пользоваться.

Допустим, что при угадывании первой буквы слова, следующего за  $w_1$ , испытуемые (30 человек) по данным протокола записали следующие буквы: а - 15 человек, д - 8, в - 2, б - 2, с - 2, р - 1. Двадцать букв французского алфавита не были названы. Исходя из этих данных, производим следующий расчет: каждому из двадцати неназванных продолжений приписываем вероятность  $p_h = 0,01$ , отсюда  $p_h$  двадцати букв равняется  $p_h = 0,01 \times 20 = 0,20$ , а сумма вероятностей названных продолжений равна  $1 - 0,20 = 0,80$ ; б) высчитываем вероятность каждой названной буквы:

$$p_a = \frac{15 \times 0,80}{30} = 0,40, \quad p_d = \frac{8 \times 0,80}{30} = 0,213,$$

$$p_v = p_b = p_c = \frac{2 \times 0,80}{30} = 0,053, \quad p_p = \frac{1 \times 0,80}{30} = 0,027;$$

в) подставляем значение вероятностей<sup>ж</sup> каждой буквы в формулу I, например:  $I_a = H_a = 0,40 \log_2 0,40$ ,  $I_d = H_d = 0,213 \log_2 0,213$  и т.д. и получаем оценки I для каждой названной буквы:  $I_a = 0,5288$ ,  $I_d = 0,4752$ ,  $I_b = I_r = I_c = 0,2246$ ,  $I_p = 0,1407$  двоичных единиц (дв. ед.); г) вычисляем энтропию (информацию) неназванных букв

$H = 0,01 \log_2 0,01 \times 20 = 0,0664 \times 20 = 1,3280$  дв. ед.; д) суммируем все оценки энтропии и получаем величину I, несомой угадываемой буквой текста.

$$I = 0,5288 + 0,4752 + 0,2246 \times 3 + 0,1407 + 1,3280 = 3,1415 \text{ дв. ед.}$$

Аналогичным образом обрабатываются результаты угадывания остальных букв диагностирующего сегмента.

Следует отметить, что величина I в значительной мере зависит от количества возможных продолжений на каждом шаге текста. Поэтому остановимся на этой процедуре.

Возьмем фразу:

Г	е	Δ	g	о	у	в	е	г	п	е	ш	е	п	t	Δ	а	Δ
I	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
а	п	п	о	п	с	é	Δ	l	а	Δ	d	é	v	а	l		
19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34		
и	а	t	i	о	п	Δ	д	и	Δ	f	г	а	п	с	Δ		
35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50		

Поскольку угадывание текста начинается с первой буквы слова, следующего за  $w_1$ , то и определение количества возможных продолжений начинается с той же буквы, т.е. с семнадцатого шага в нашем примере.

Итак,  $w_1$  известно. Это существительное, относящееся к общественно-политической лексике. Учитывая это, необходимо найти на каждую букву французского алфавита хотя бы по одному слову, которое по своей лексико-грамматической сочетаемости могло бы следовать за  $w_1$ . Такими словами, прежде всего, должны быть глаголы и прилагательные. Обращаясь к словарю, устанавливаем, что на данном шаге возможны все буквы, кроме W и X. Значит, следует искать другие части речи. Оказывается, что в газетных текстах достаточно широко распространены

<sup>ж</sup> Табличные величины  $p_i \log_2 p_i$  можно найти в книгах: И.М. Яглом и И.М. Яглом. Вероятность и информация. М., 1967, стр. 305-308. "Статистика текста". Л., 1968, стр. 231-258.



словосочетания типа *le gouvernement Willy Brandt*. Значит, буква *W* на семнадцатом шаге возможна. Что касается буквы *X*, то ее нужно исключить из числа возможных продолжений. Таким образом, для данного шага фиксируем двадцать пять продолжений.

Рассмотрим восемнадцатый шаг. Известны  $W_1$  и первая буква второго слова. Принцип поиска остается тот же, но диапазон возможных продолжений сужается — теперь подбираются только слова, начинающиеся на букву *a*. В соответствии с комбинаторикой на этом шаге были бы возможны двадцать четыре буквы и знак пробела  $\Delta$ .

Лексико-грамматический анализ показывает, что слова, начинающиеся на *ab*, *ah*, *ak*, *ao*, *aq* не могут следовать за  $W_1$ . Поэтому считаем, что на данном шаге возможны двадцать продолжений.

Следует помнить и об экстралингвистических ограничениях. Так, например, на сорок пятом шаге возможны только те буквы, на которые начинаются слова, обозначающие денежные единицы. Это диктуется словом *la dévaluation*. Однако, поскольку  $W_1$  не указывает, о правительстве какой страны идет речь, необходимо иметь в виду денежные единицы, о курсе которых сообщается в газетах Международным банком.

Для обработки результатов второго варианта угадывания поиск продолжений ведется аналогичным образом. Но если сравнить количество возможных продолжений на одном и том же шаге в обоих вариантах, то во втором случае их будет больше. Так, в нашей фразе на семнадцатом шаге во втором варианте возможны все двадцать шесть букв алфавита, а на восемнадцатом — двадцать четыре. Объясняется это тем, что в связи с отсутствием  $W_1$  на комбинаторику букв не накладываются лексико-грамматические (контекстные) ограничения.

#### Отбор лингвистического материала

Объектом наших семантических измерений является лексика газетных текстов, т.е. текстов, относящихся к публицистическому стилю.

Подбор лексики осуществлен по частотному принципу. В соответствии с этим принципом наша лексика делится на две группы:

I. Лексика, имеющая в словаре И.А. Исенина (1968) высо-

кую частоту: *gouvernement* m (116), *peuple* m (88), *classe* f (60), *état* m (57), *ordre* m (50), *président* m (41), *exploitation* f (39), *défense* f (36), *ministre* m (35), *régime* m (34).

2. Лексика, имеющая в том же словаре низкую частоту, — менее 10: *armistice* m, *sécession* s, *délégation* d, *moratoire* m.

Полисемантические слова исследовались в нескольких значениях. Последние указаны в таблице I вместе с частотами их появления в выборке объемов в I млн. 80 тыс. словоупотреблений. При обсуждении полученных результатов каждое значение рассматривалось как отдельная лексическая единица. Таким образом, общее количество единиц, подвергнутых анализу, составляет 30.

Как показывает перечень слов, мы измеряем смысловую информацию только существительных. Выбор части речи в качестве объекта исследования вытекает из основного условия измерения смысловой информации в нашем эксперименте, согласно которому контрольное слово не должно подвергаться влиянию контекста, т.е. оно должно занимать начальную позицию в предложении. Во французском предложении в этой позиции, как правило, находится существительное или личное местоимение, выполняющие функцию подлежащего.

### Обсуждение результатов эксперимента

34 опыта (под опытом понимается двойное угадывание текста, т.е. с контрольным словом и без него), в которых приняло участие 400 учащихся и учителей одного из парижских лицеев, дали 34 оценки Исм.

Анализ полученных данных позволил установить, что Исм слова в опытах с носителями языка зависит от лингвистического тезауруса информантов, частотности и значения слова, лингвистических и статистических характеристик диагностирующего сегмента (правого контекста) и пр. (Богодист, 1974).

Эти наблюдения представляют значительный интерес для лингвистики.

Для методики преподавания французского языка в вузе было бы целесообразно провести аналогичный эксперимент со студентами разных курсов и проанализировать оценки Исм, сопоставив их с оценками Исм, полученной в опытах с носителями языка.

136 опытов, осуществленных с этой целью в 10 вузах нашей

страны, в которых приняло участие 400 студентов отделений французского языка, дали 136 оценок Исм (см. табл. I).

Анализ этих данных можно резюмировать следующим образом.

1. Описанный метод является вполне корректным инструментом измерения Исм слова не только в текстах, воспринимаемых носителями языка, но и в текстах, воспринимаемых иностранцами, в различной степени владеющими данным языком.

2. Метод угадывания в том виде, в каком он применялся нами, позволяет получить количественные оценки, характеризующие степень усвоения студентами разных курсов не только структуры французского языка, но и его семантической системы.

Оказывается, что информативность французского слова в опытах со студентами всех курсов ниже, чем в опытах с носителями языка.

Например, средняя Исм слова для французских учащихся и студентов IV-У, III, II и I курсов соответственно составляет 7,60; 7,00; 6,46; 5,98 и 4,10 дв. ед.<sup>\*</sup>

3. Оба эксперимента свидетельствуют о наличии связи между Исм слова и правым контекстом. Эта связь проявляется в различной степени в зависимости от уровня владения языком - в опытах с носителями языка для извлечения максимальной Исм, несомой контрольным словом, достаточен контекст из 5-6 слов, в опытах со студентами IV-У курсов - из 6-7, III курса - из 6-8, I-II курсов - из 8-9 слов.

4. Исм слова варьируется в зависимости от частотности слов, составляющих диагностирующий текст. Так, в текстах, в которых средняя частота составляющих их слов (политическая лексика) находится в интервалах от 4000 до 3000, от 3000 до 2001, от 2000 до 1001, от 1000 до I, средняя Исм контрольных слов для носителей языка составляет 9,48; 8,86; 7,80 и 6,32 дв. ед., для студентов IV-У курсов - 7,12; 7,08; 6,50; 7,98; 6,40 дв. ед.; III курса - 6,72; 6,64; 6,22; 5,44; 6,60 дв. ед.; II курса - 4,35; 5,38; 5,30; 5,00; 5,41 дв. ед.; I-го

\* Чтобы вычислить среднюю Исм, нужно суммировать оценки Исм отдельно по каждой колонке таблицы I и разделить на 34. Для IV и У курсов в таблице имеется только одна колонка. Объясняется это тем, что результаты эксперимента с этими студентами были объединены после того, как оказалось, что величины Исм в обоих случаях близки.

курса - 3,79; 4,99; 4,46; 4,24; 4,65 дв. ед.<sup>ж</sup>

Таким образом, как для носителя языка, так и для иностранцев, в достаточной степени владеющих данным языком, т.е. студентов III, IV, V курсов, смысловая нагрузка контрольного слова выше в текстах, состоящих из более частых слов. Что касается оценок Исм, полученных в опытах со студентами I-II курсов, то здесь такая зависимость не наблюдается.

5. В опытах с французами установлено, что Исм слова зависит от лексической сочетаемости слова. Этот факт находит свое подтверждение и в опытах со студентами всех курсов, кроме первого. Это означает, что студенты усвоили не только слова, но и их лексическое окружение.

6. Как в опытах с носителями языка, так и в опытах со студентами вузов нашей страны для разных значений слов получены различные оценки Исм.

В первом случае установлена следующая закономерность: более редкие значения слов несут больше Исм.

Тот же результат дало угадывание текстов студентами IV-V курсов.

Опыты со студентами III, II и I курсов свидетельствуют о том, что здесь этот фактор менее релевантен, а в ряде случаев более информативными оказались более частые значения слов.

Приведенный выше анализ результатов эксперимента свидетельствует о том, что описанный метод измерения Исм слова позволяет получать количественные оценки, характеризующие уровень владения студентами разных курсов семантикой французского языка, что открывает перед методикой обучения этому языку в вузе новые возможности.

Так, например, можно измерить Исм синонимов, омонимов, антонимов, членов предложения, частей речи, грамматических конструкций и пр. для носителей языка и для студентов разных курсов, изучающих данный язык, и в соответствии с полученными оценками Исм разработать более гибкую методику обучения иностранному языку.

Можно также поставить задачу измерения Исм контрольного

---

<sup>ж</sup> Средняя частота слов, составляющих диагностирующие тексты, получена следующим образом. Для всех слов по словарю G. Gougenheim, P. Rivenc, R. Michéa, A. Sauvageot, L'élaboration du français fondamental (P., 1964) были найдены частоты, а затем вычислено их среднее арифметическое.

слова в словосочетании. В последнем случае, представляющем большой интерес и для лингвистики, и для методики, процедура проведения опытов и математическая обработка являются менее трудоемкими.

В настоящее время описанный метод может использоваться в качестве контрольного семантического теста, однако, если бы были проведены более широкие семантические измерения текстов и составляющих их лингвистических единиц, то отбор лексического и грамматического материала мог бы осуществляться на информационной основе.

### Литература

- Богодист В.И. Измерение смысловой информации лингвистических единиц французского текста. Автореф. канд. дисс. Л., 1974, стр. 14-16.
- Богодист В.И., Георгиев А.Ц., Пестунова В.Н., Пиотровский Р.Г., Райтар С.В. Как измерить смысловую информацию. - ЦИВШ "Филологические науки", 1975. № 4, стр. 91-101.
- Исенин И.А. О частотном словаре подъязыка современной французской прессы. - В сб.: Статистика речи, Л., 1968, стр. 185-190.
- Пиотровский Р.Г. Измерение семантической информации. - В сб.: Психологические механизмы памяти и ее закономерности в процессе обучения. Харьков, 1970, стр. 117.
- Шеннон К. Математическая теория связи. - "Работы по теории информации и кибернетике". Перев. с англ. М., 1963, стр. 394.
- Bogodist, V., Georgiev, Ch., Pestunova, V., Piotrovski, R., Raitar, S., Semantische Information und Arten ihrer Messung. "Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung", 1975, Band 28, Heft 2, S. 221-235.

Семантно-прагматическая (семоловая) информация французских слов

Таблица I

Лексические единицы. Значение.		Семантно-прагматическая информация в лексич. единицах						
		абс.	фр.	IV-V курс	II курс	II курс	I курс	
1	2	3	4	5	6	7	8	
1. exploitation f	"разработка"	9	12,03	10,85	8,84	8,01	7,02	
2. "	"эксплуатация"							
3. "	"угнетение"	56	9,59	9,41	8,88	8,00	7,98	
4. "	"участок земли"	26	6,27	4,66	4,46	4,16	2,46	
5. classe f	"класс (социальный)"	116	9,24	7,62	7,61	7,46	6,84	
6. "	"классная комната"	5	9,69	8,87	8,61	7,00	6,44	
7. armistice m	"перемирие"	2	7,43	6,54	6,20	4,67	3,81	
8. sécession f	"отпадение, отделение"	2	10,60	9,06	9,00	5,81	5,68	
9. régime m	"режим (река)"	2	10,31	6,30	6,00	5,80	5,00	
10. "	"гос. строй"	3	9,24	8,00	5,00	3,72	3,42	
11. "	"режим (мед. терапия)"	128	6,67	5,96	5,75	5,29	5,15	
12. moratoire m	"мораторий"	4	5,27	5,55	5,34	3,77	2,20	
13. défense f	"защита (военная)"	2	9,77	7,98	5,99	5,31	3,33	
14. "	"защита (правическая)"	19	7,02	3,63	1,87	1,17	1,52	
15. "	"защит"	6	9,48	8,89	7,28	7,01	6,47	
16. état m	"состояние"	2	9,55	8,31	7,66	5,13	4,66	
17. "	"государство" <sup>1</sup>	17	9,52	9,00	8,75	7,98	7,24	
18. "	"государство" <sup>2</sup>	179	7,02	6,83	6,62	6,18	4,11	
19. président m	"президент" <sup>1</sup>	179	7,68	5,91	5,71	5,40	4,11	
20. "	"президент" <sup>2</sup>	123	7,59	6,62	5,23	2,62	2,00	
21. "	"председатель"	123	7,29	7,24	5,54	5,29	4,18	
22. gouvernement m	"тип. гос. управления"	34	5,15	4,91	4,00	2,82	2,01	
23. "	"правительство"	2	5,92	5,20	5,03	4,81	4,00	
24. "	"управление"	491	5,44	5,56	4,56	4,00	1,94	
25. peuple m	"народ"	2	4,68	7,69	5,55	5,18	3,98	
26. "	"народность"	144	6,63	5,50	5,50	5,50	4,50	
27. "	"племя"	8	8,78	7,99	7,88	6,10	4,20	
28. ministre m	"министр" <sup>1</sup>	4	5,83	6,43	5,41	4,38	4,19	
29. "	"министр" <sup>2</sup>	270	5,87	5,44	5,31	4,00	3,87	
30. "	"министр" <sup>3</sup>	270	6,30	6,13	5,64	4,45	3,87	
31. ordre m	"порядок"	270	5,95	5,68	5,63	5,35	5,13	
32. "	"порядок (в квартире)"	14	6,07	5,62	5,11	5,11	2,00	
33. "	"приказ"	2	7,34	7,97	7,11	6,12	6,00	
34. délégation f	"делегация"	12	6,68	6,77	6,15	5,26	5,00	
		125	6,82	6,50	6,13	5,13	4,75	

Примечание: Цифры 1, 2, 3, которыми отмечены некоторые слова в столбике 2, указывают номер эксперимента.

INFORMATION SÉMASIOLOGIQUE VÉHICULÉE PAR UN MOT  
AUX SUJETS PARLANTS ET AUX ÉTRANGERS

V. Bogodiste

R é s u m é

L'article décrit une méthode psycholinguistique concernant l'évaluation de l'information sémasiologique véhiculée par un mot français aux Français et aux étrangers (Russes). On constate que l'importance de l'information sémasiologique du mot dépend de thésaurus du sujet, de la fréquence du mot, de sa signification etc. Ces conclusions ne manquent pas d'intérêt pour la linguistique et la didactique du français.

## THE ENGLISH ORAL TEXT: A TYPOLOGICAL STUDY

B. Gasparov, V. Paperno

I. Studies of Russian oral texts which began nearly fifteen years ago have shown that speech - or the oral text - is not simply a reduced and distorted version of standard written language. Rather, the oral text is a language mechanism which has its own specific structure and conforms to its own laws (Земская, 1973, 18-25). The specific structure of oral texts manifests itself in expressly "oral" phrases, grammatical constructions, and so on, but it also manifests itself at a higher level - by specifically oral means of unfolding a speech sequence and expressing its message. The basic composition of the written text is a linear structure: the sequence of thematic units is controlled by a set of rules, and the composition of the whole text is determined by a rigid order of units, for each thematic unit has its own function which determines its position in the entire structure. Oral communication calls for a different structure, for a speaker can not compose his speech according to a complicated set of rules. By its very nature, oral communication does not tolerate rigid linear structures. Moreover, a speaker enjoys means of expression which a writer does not have: melodic and rhythmical devices, gestures and facial expression. The speaker continuously employs these means, and with their help creates a non-linear structure - a text which unfolds simultaneously along several channels, as it were. Therefore, the chief typological feature that distinguishes the oral from the written text is its non-linear character.

The grammatical relationships within the oral text do not display the rigidity characteristic of the written text. The basic units of the oral text tend to be superimposed upon each other, to add to each other rather than to be connected in accordance with grammar rules. The message of the oral text does not, therefore, unfold in a logical sequence. It is the sum total of its elements. Oral and written texts are the realization of two different sublanguages, each reflecting a certain type of language mental activity and realizing its own function in the cultural mechanism.



The non-linear character of the oral text entails numerous breaks in grammatical structure. The apparent "ungrammaticality" of speech may be ascribed to the inability of the speaker to control every aspect of his text at the pace of its creation. The grammatical breaks, however, have their own positive function, and the seemingly "disorderly" speech has, accordingly, its positive value: the non-linear, multi-layered, polyphonic oral text fully realizes the possibilities of oral communication.

These observations can be amply illustrated with quotations from Russian oral texts. One must remember, though, that certain aspects of Russian grammar (comparatively free word order, etc.) make it easy for the Russian speaker to create a non-linear structure. In Russian it is easier to break a phrase or to insert an aside into a grammatically complete phrase than, for example, in English. Standard English grammar, with its strict word order, its complicated system of tenses, and its articles requires comparatively more rigid structures and therefore is more strongly oriented towards linear texts. Consequently, it is more interesting to reveal the non-linear character of the English oral text and to see how the tendency towards non-linear structures in oral communication manifests itself in a language strongly oriented towards linear composition.

2. First, let us consider<sup>8</sup> the general structure of the oral text. The following text is an excerpt from a public talk.

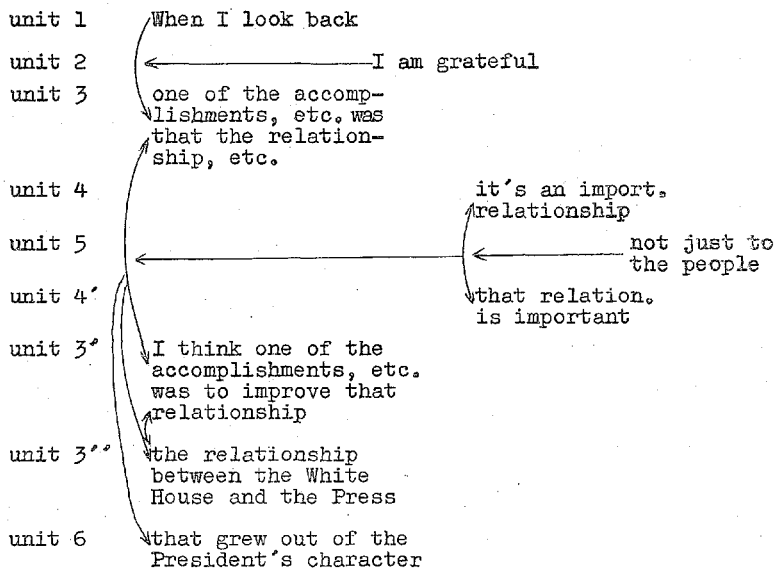
"When I look back though, uh, and I am grateful for...<sup>8</sup> for ... fo ... to Martha for mentioning it uh ... one of the accomplishments of the Ford Administration was that the relationship was uh ... between the Press and the White House, and it's an important relationship ... uh, not just to the people who live and work in Washington but really to all Americans whose view of reality ... uh is what the Press uh says is uh reality, so that relationship is important, and I ... I think one of the accomplishments of the Ford Administration was to greatly improve that relationship, to re-

---

<sup>8</sup> This is the way we designate a pause; we do not omit any part of our samples. All the samples represent authentic contemporary American speech and were extracted from recordings of radio broadcasts and of conversations in typical social situations.

store a good deal of civility, of trust, uh ... friendship even uh ... to the relationship between the White House and the Press, and that largely grew out of the President's own traits of character as well as 25 years in Washington in which he had uh ... he had come to know and like and become uh really personal friends with many reporters ..."

The text consists of a number of thematic units. Each unit is a grammatically independent utterance with a theme of its own; wherever the grammatical connection is broken, i.e. in the interval between the end of one unit and the beginning of the next, we find a filler, vocal or silent (more about fillers later). The thematic unfolding of the text is not achieved by a direct logical sequence of units; rather, as the theme develops, certain units - or their parts - keep reappearing or repeating, and thus form thematic circles, as it were. The units within the circles can be compared to parentheses or asides. The reappearance of a theme, its variations and repetitions create a fragmented, but fuller and more complex representation of the speaker's idea; they create the oral text. The listener comprehends the speech not, again, as a direct logical sequence of thematic units, but as a whole, as the sum total of all the interacting variations and repetitions.



Unit 1 in the above text and diagram serves as an introduction; before proceeding to the main theme, however, the speaker inserts an aside - unit 2; then unit 3 brings us back to the main theme which once again is stalled by a digression - unit 4, later duplicated by unit 4'. Between these two units (4 and 4'), inside the "circle", the speaker allows himself another aside - unit 5, a commentary on a commentary. Only after this does the speaker proceed with his theme - units 3' and 3''. It is worth noting here that units 3, 3', and 3'', which develop the main theme, are not simple repetitions: at each stage new information is added to the theme. Taken together, all these variations, repetitions, and asides create a general description of the relationship between the White House and the Press, and it is to this description as a whole that unit 6, the conclusion, is attached.

An even more complicated text structure will be observed in informal speech. Let us watch this "snowball" growth of description in a story told in a private conversation.

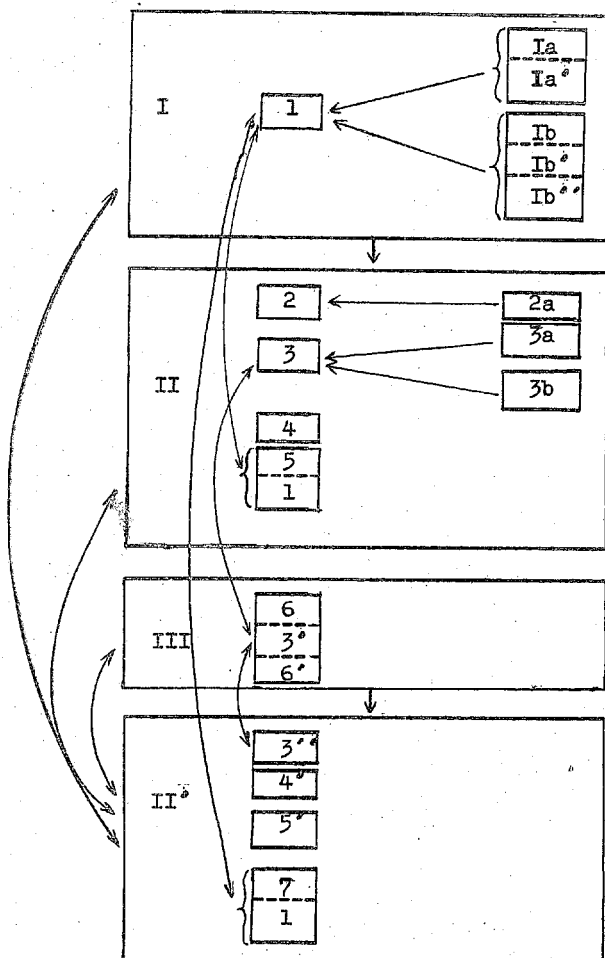
I<sup>1</sup> "His mother was paranoid about anybody knowing that he'd gone to China,<sup>1a</sup> in fact she told all her relatives that he'd gone to Africa to work,<sup>1a'</sup> she didn't even tell them where he was,<sup>1b</sup> and she didn't want him to write,<sup>1b'</sup> she was afraid to get mail from him,<sup>1b''</sup> she was afraid to communicate at all ... and so ...<sup>2</sup> um, when Jane Brown was here,<sup>2a</sup> you know who that is, don't you? I told you about her ...<sup>3</sup> she came to the dorm and brought me some stuff, some presents that he had sent for his mother and his sister, and a picture of him ... um, and a letter, and Chinese clothes for them, and some stuff, souvenirs and stuff, which I took over there ...<sup>4</sup> and his mother op ... you know, opened the stuff, and saw the picture of him, and she just started crying,<sup>5</sup> it was the first time she had heard from him in something like a year and a half or two years ...<sup>6</sup> because she was just afraid to communicate ... and ...<sup>6</sup> and that ... that was really ... that was very strange because I didn't know this woman at all, and I hardly knew her son,<sup>3'</sup> and here I was bringing ...<sup>6'</sup> I didn't rea ... I didn't really even talk to him<sup>3''</sup> ... and here I was bringing all that stuff,<sup>4'</sup> and she's just sort of cry-

ing,<sup>5</sup> you know, because she hadn't heard anything from him for so long, and ...<sup>7</sup> and even when ... she made it clear to me that if anybody came over when I was there I wasn't to say a thing about where he was,<sup>1</sup> that everybody is supposed to think he was in Africa."

This text also can be easily segmented into thematic units (1 - 7) separated from each other by fillers which mark breaks in grammatical structure. Some of these units are continuously duplicated in the text and form a complicated system of thematic interweaving and superimposing. In addition, units 1,2,3 are subdivided into a number of sub-units which also enter into a complex relationship; sub-unit Ia, for instance, appears in the text three times, each time with new variations. Besides, units 1 - 7 can be easily grouped into super-units. I - Introduction, II - Story, III - Summary, and II' - Appendix, a variant of super-unit II; super-unit II' repeats the story and adds new information.

The super-units are related to each other in the same way as the units and the sub-units, i.e. they partly repeat each other, partly add to each other, and together create the snowball, non-linear, polyphonic structure of the oral text. Typical of its polyphonic character is the reappearance of super-unit II as a variant, i.e. super-unit II' ; although the story has been brought to a close in the Summary (super-unit III), the speaker does not stop, but adds an appendix, super-unit II' which renarrates the story, supplements it with variations and additions (unit 7), and - since it follows the Summary - sheds new light on the story as a whole: "That was very strange, etc." Note also the reappearance of unit I at the end of super-units II and II': unit I recounts the theme of the Introduction. (See the scheme on p. 20).

3. In creating the multi-layer, polyphonic oral text, the speaker uses certain specific devices; the most important of these are 1) fillers, 2) word order specific for oral texts, 3) articles specific for oral texts, and 4) specific usage of tenses. These devices form the plane of expression that reflects the non-linear structure of the plane of meaning.



### 3.1. Fillers.

Oral text, both formal and informal, is always segmented into word groups separated by what we call fillers. The simplest of these is a pause, sometimes accompanied by a quasi-interjection (uh, um, mm, etc.)<sup>32</sup>. Certain words or

<sup>32</sup> See the review of the works on the problem of "hesitations": (Николаева, 1970).

phrases can also serve as fillers, and in this function they lose all lexical meaning. Some of the most frequently used fillers of this type are and, so, well, O.K., you see, you know, I mean. The fact that even and loses its lexical meaning when used as a filler may be illustrated by the following example: "I got up at 9:30 in the morning ... which I want you to know was an effort on my part, and ... I got up at 9:30 in the morning ... I hid in the stairwell ..." Some of the words and phrases used as fillers seem only partly to lose their lexical meaning. Such fillers as kind of, sort of, like continue to express the idea of approximation or similarity even when their usage is clearly that of a filler. E.g., "... for about two hours ... I don't know ... well, like, you know, really a long time." Finally, another type of filler is a mere repetition of a word or several words. Note the frequent use of these types of fillers (in several instances we even encounter a whole string of fillers) in the following monolog.

"I heard a funny story about some uh guy that bought a cheap tape recorder ... and it had a warranty on it, you know ... well, he took it home ... and he turned it on, it worked maybe ten minutes ... and it burned out ... stopped. So he went back ... to the ... uh ... to the store, and he said, Look, my tape recorder burned out on me, you know. So, uh ... says, Well, O.K., O.K., if ... leave it here, you know, we'll take a look at it. So he came back two weeks later ... nothing was done.... hadn't been touched ... so he said, Well ... look, you know ... what am I gonna do? I mean I bought this tape recorder, it doesn't work ... well, they said, Look, you know, the thing is we can't ... we don't have many parts and ... but ... but if it's ... if it's an emergency repair you can fill out these special forms ... and we can try and get to it, you know. So he does, he fills out all these forms and ... I mean, he had to take it round to get a signature or something ... from somebody, you know ... and, Come back! And so he brings it back and ... O.K. ... leaves it ... goes back two weeks later. Well, they ... they've taken a look at it ... but they said, Well, you know ... uh ... there's some malfunction here ... in the motor or something and ... uh, you have to take it back to the factory to get it

fixed because we just don't have the ... the part, and ... nothing we can do. So O.K., a month has gone by ... so he takes it to the factory ... and uh ... and he ... he says, Look, I need this part, and they won't give it to me in the store, and ..."

In a dialog fillers are used just as often; the interesting difference is that sometimes when a speaker needs a filler, it is supplied by the other speaker, e.g., "I am sure if Betty started sleeping there, you know ... / Yeah, on that bed ... / she would have said something sooner or later." The second speaker supplied a filler /Yeah/ after the first speaker had already used one (you know).

The position of a filler in the text is not determined by the syntactical structure of the text (cf. Launsbury, 1954, 98); fillers are found at the beginning of a phrase, at its end, and even within a syntactically complete phrase. This means that fillers have nothing to do with syntactical segmentation of the oral text; fillers are not "oral punctuation marks", as might be assumed. The function of the filler is closely connected with the multi-layer structure of the oral text.

The fact is that a filler which follows an utterance cancels the semantic and grammatical obligations imposed on the speaker by what he has already said. Having inserted a filler, the speaker feels free to start a new syntactical sequence (and to introduce a new idea) regardless of whether he has completed his previous statement or not. And so the speaker has a chance to correct or complete his statement; e.g., "When I look back though, uh, and I am grateful for ... for ... fo ... to Martha for mentioning it ..." "... and so she said, Oh well, well, we ... it was graduation day ... she said, We just came up to see our son graduate." "And you want to do things that other people do ... who are your age ... and you're uh, a long time fussing with it ... and you can't do it ... it must be ... I mean ... absolutely shattering." "... and he was sort of bald and kind of... you know, had a kind of pot ... and ..."

So the primary function of the filler is to enable the speaker to repeat his message at any point, to add to it, to introduce asides and parentheses, to combine and juxtapose

semantic units in a nonpredetermined order, and thus create a multi-layered, non-linear, polyphonic, snowball representation of his story. Note, for example, the zigzag of the theme in the following dialog:

"Yeah, I really should try to come back before I go home again ... / Yeah, that's why I stayed, you know, as long as I did ... you mean, from Europe? / Yeah, I was ... I'll feel better ... I won't mind doing it and I won't mind spending the money ... / Uh-uh. / ... if I have a job waiting for me."

Note also the variations of the theme in this exchange: "She's really doing a job in there ... you know, I'm really impressed with her house cleaning ... / Yeah, she's really ... / thorough."

### 3.2. Word order.

Studies of Russian oral texts have shown that in Russian a specific "oral" word order is one of the most effective devices for creating a non-linear text. Its most striking peculiarity is that the phrase tends to be disjointed, its configuration broken, and its components found sometimes at the opposite ends of the utterance; e.g., "Книжку я вчера прочитал интересную очень". The standard word order would be: "Я вчера прочитал одну очень интересную книжку". As a result of this broken word order, an utterance has the same non-linear structure as the whole text.

The rigid word order of standard English does not allow the English speaker as much leeway as the Russian speaker. The occasional violations of the standard English word order which we observe in English oral texts are comparatively rare and less obvious; e.g., "When are you gonna give me back that shirt? You're gonna give it to me back dirty I suppose." "Oh, I saw the first Pole I have ever seen ... in our dorm last night." "I showed you that list in ... in the newspaper of people that passed the Bar?" "Instead of having to close the store for inventory they ... every day know exactly what's gone out of the store, what's come into the store ..." The pauses (fillers) in some of these utterances probably indicate that we are dealing here with a borderline case between irregular word order and the "zigzag" of the theme. These



borderline cases may be taken as an illustration of the fact that the rigidity of English word order entails comparatively more frequent use of fillers in the English oral text and comparatively shorter segments that are syntactically correct.

There exists, however, one case of a specific "oral" word order that is characteristic for both English and Russian speech. It occurs when a word or a word group which is thematic for the utterance takes the position at its opening - or closing - while its standard position may or may not be filled by a substituting pronoun; e.g., "I mean, why should that make any difference, wooden beds?" "Yeah, that's the last thing they'll get, 'cause that's the last thing anyone does in that dorm, study." "But the other museum they stayed in for an hour ..." "Actually, the grammar they can even do in bigger groups." (cf. Labov, 1975, 14). When combined with a filler, this device borders, again, on a zigzag of the theme, e.g., "Where did you get that sweater? It's pretty. / It's Norwegian ... My ... my parents ... they bought it for my father although it's a little too small for him." "In my other family, my Uncle ... his wife died about five years ago, and ... uh ... they both ... and he and this other woman are very proper society people and everything, and they live together."

The close connection between these two devices - the filler and the irregular word order - points to their common function in the composition of the non-linear, zigzag oral text.

### 3.3 Quasi-article.

In the written English text, one of the functions of the indefinite article is to mark the first appearance of a noun; when it reappears in the same text it is preceded by the definite article. Thus the pair a - the helps to establish the linear character of the written text. And it is precisely this linear effect of the pair that makes it very rare in the oral text. Instead, the speaker tends to use quasi-articles that produce no linear effect. Quasi-articles are semantically transformed demonstrative pronouns this/these and that/those; e.g., "It was about a year ago, I

think it was, I had this apartment in San Francisco, I lived on the third floor of this little tiny apartment building, I had a little two rooms and a dog, right? I live up on the third floor, and I used to walk around town and ... I had, you know, a couple of pairs of levis and ... a coupla T-shirts and I thought I had my shit together pretty good man, you know. I was out on the streets ... talk and I talk, and all that shit, and every time I found a nice piece of town. He went right straight down the stairs to this chick on the second floor, right?" In "this apartment" and "this building" the pronoun obviously has no demonstrative meaning since neither the apartment nor the building is present, near or previously mentioned; in a written text the indefinite article would normally be used in such a situation. The same applies to "this chick", of whom the listener has no previous knowledge (note the explanation added by the speaker, "there was another chick").

The article function of "this" can be even more clearly demonstrated by the following quotations: "The other bed had this ... had a ... /Spring?/ ... spring that came out ..." "He called them ... and uh ... apparently this ... the girl there that's kind of ... in M&B ... had been leaving notes for John." Here the speakers hesitate between using a or this and the or this, and it is significant that the quasi-article competes with both the indefinite and the definite article, for it shows again that whether or not the noun was mentioned earlier in the text is irrelevant as regards quasi-articles. Thus the quasi-article carries out the semantic functions of the article without making the oral text linear. In contemporary speech quasi-articles are fairly frequent. Here are a few more examples: "Well, you know, one problem they have in their economic system is ... uh, is ... is this expectation of ... of ... continual ... uh ... productivity raises." (Note the use of this with an abstract noun) "So in the next planning period they'll always be able to improve ... a little bit, and thereby get all these bonuses." "I saw this thing on Fifty Minutes Fun ... and he says, Well, sure, everybody is talking about young people, how they live together now, and ... but uh really, all the old people li ... it's really funny, you know ... like maybe this ... this old

guy's wife's died, and this old lady ..." (Note that here "this old guy" and "this old lady" do not denote any specific person; this is a case where normally "some" would be used) "Someone knocked at the door, and I answered the door, and this lady with this accent said to me, Is Dim home?"

The pronoun that/those seems also to be used as a quasi-article, though it is less frequent and probably retains to some extent its demonstrative meaning; e.g., "It was either a week or ten days in the Soviet Union leaving from Los Angeles, and it wasn't just Moscow and Leningrad, it was all those other places too, Tallin, I forget all that was on it." "I showed you that list in ... in the newspaper of people that passed the Bar?"

Although standard Russian does not know articles, quasi-articles do exist in oral Russian texts:

The absence of articles, together with a loose word order, has to do with the fact that even in standard Russian the tendency towards linear structures is not nearly as strong as in standard English. But even when the article - or, rather, quasi-article - appears in Russian speech, it does not have the linear effect.

The corresponding usage of quasi-articles in Russian and English speech makes their role in the oral text structure all the more obvious.

#### 3.4. Tenses.

The complex tense system of English enables the English speaker to render an accurate description of the time-sequence of events; but an accurate description of this kind can not fail to produce a linear text. And so in an oral text we seldom find a complicated system of tenses; the speaker feels that building such a system requires effort and time, and is not really necessary. Stories with a complex and significant time-relationship tend to be simplified as regards tenses; the speaker finds other ways to express his message. See for instance this story about a group of tourists who failed to obtain the necessary visas at the right time:

"It so happened that they finally did get their Samarkand visa ... I forgot from where ... but it was for after

the time ... well, for the t ... the time ... let me see ... they had to be back, let's say by eighth ... and the vi ... in Leningrad. ... and the visa for Samarkand was, uh, you know, eighth to thirteenth or something like that."

Statistically speaking, we find that, as a rule, the tense system in an English oral text is considerably simplified: a) the Perfect Continuous Tenses are used very seldom, b) the Past Perfect Tense tends to be replaced by Past Indefinite or even by Present Perfect (usually with a filler). E.g., "Oh, I saw the first Pole I have ever seen ... in our dorm last night." "He said that she even hasn't answered some kind of invitations that she's had." c) the Present Perfect Tense also tends to be replaced by Past Indefinite, and so the Indefinite Tenses dominate in any oral text. See for example this monolog:

"I heard a funny story about some uh guy that bought a cheap tape recorder ... and it had a warranty on it, you know ... well he took it home ... and he turned it on, it worked maybe ten minutes ... and it burned out ... stopped. So he went back ... to the, uh ... to the store, and he said, Look, my tape recorder burned out on me, you know. So, uh ... says, Well, O.K., O.K., if ... leave it here, you know, we'll take a look at it. So he came back two weeks later ... nothing was done ... hadn't been touched ..."

The simplification of tenses is not, of course, the mere result of the speaker's difficulties; it has its own positive, creative function, i.e., it creates a non-linear structure which is characteristic of an oral text: the story ceases to be a string of actions carried out in a certain order, but turns instead into a "snowball" whole, with separate events fusing together to compose a polyphonic oral text. Furthermore, to ensure the multi-layered structure of his "snowball", the story-teller uses special devices which have a non-linear effect: a) the speaker continuously "zigzags" between the present and past tense; by changing his position in time, the speaker acquires an opportunity to set different time-keys to his narration, to re-adjust his perspective, and thus add to the polyphony of the text; b) the speaker often uses a narrative device that may be defined as quasi-quotation, e.g., "... in fact there was an article that says this is against the law." "He said that she even hasn't an-

swered some kind of invitations that she's had." Strictly speaking, these are not quotations, and so in standard English a change of tenses would be unavoidable here; the speaker, however, feels that a quotation and quasi-quotation justify the conflict of tenses. The following story will illustrate the use of these tense-devices in the oral text.

"... so he said, Well ... look, you know ... what am I gonna do? I mean I bought this tape recorder, it doesn't work ... well, they said, Look, you know, the thing is we can't ... we don't have many parts and ... but ... but if it's ... if it's an emergency repair you can fill out these special forms ... and we can try and get to it, you know. So he does, he fills out all these forms and ... I mean, he had to take it round to get a signature or something ... from somebody, you know ... and, Come back! and so he brings it back and ... O.K. ... leaves it ... goes back two weeks later. Well, they ... they've taken a look at it ... but they said, Well, you know ... uh ... there's some malfunction here ... in the motor or something and ... uh, you have to take it back to the factory to get it fixed because we just don't have the ... the part and ... nothing we can do. So O.K., a month has gone by ... so he takes it to the factory... and uh ... and he ... he says, Look, I need this part, and they won't give it to me in the store and ..."

The speaker alternately uses the past and the present tenses, and introduces quotations and quasi-quotations in the past and present tenses. The verbs of speaking which introduce the quotations are used sometimes in the past tense and sometimes in the present. All this contributes to the multi-layered, polyphonic structure of the oral text.

4. Thus, certain aspects of the oral text set it apart from the standard written language. As a rule, neither the speaker nor the listener is aware that the oral sublanguage conforms to its own system of laws. Popular attention usually focuses on the written text, and it is by the written patterns that our idea of the language is formed. Actual speech is rarely apprehended as it is; rather, it filters through preconceptions of what it should be. When the attention of a speaker (English or Russian) is turned to characteristics of his own oral text, he is often shocked by

what he considers to be ungrammaticality or incoherence. This naive attitude towards the oral language is exemplified by the reaction of those who read the famous transcripts of the White House tapes: "Burns, the President's chief domestic aide and counselor during the first two years of his administration, was especially shocked. Nixon had spoken clearly, decisively, grammatically and even eloquently to him. He had rarely sworn beyond a mild 'goddam'. The transcripts were full of Presidential grunts, unfinished sentences, meandering thoughts which began almost nowhere and finished someplace else. Burns had never heard some of the expressions. He thought, 'Here is a Dr. Jekyll. A split personality. What does it all mean? Does Nixon live a double life? Is he a great actor?'" (The Final Days by Bob Woodward and Carl Bernstein, Simon and Schuster, New York, 1976, p. 168) The fact is, however, that where language is concerned we all lead double lives: we read and write linear texts, but speak and listen in a non-linear way. The authors of this essay tend to think that Mr. Burns had indeed heard the "grunts, unfinished sentences, and so on." However, he, like any other listener, subconsciously "edited" the oral texts, and so far as grammar and style are concerned he heard what he expected to hear, i.e. he appreciated Nixon's speech because he heard it. The "shock" came when he saw the transcripts, and this has to do with the fact that the White House transcripts were not "edited".

In spite of the speaker's ignorance of the specific laws of the oral text, he possesses a certain subconscious knowledge of them, regularly puts them to use, composes his speech in accordance with them, and creates a multi-layered, polyphonic oral text which is comprehended - subconsciously - by the listener in accordance with the same laws.

Oral communication imposes specific demands on the participants. Hence the special structure of the oral text, a structure so powerful that even with languages as different as English and Russian it is possible to discern rules that are common to both. The difference between the grammatical structures of the English and Russian written text is considerable; the difference between the oral texts is noticeably smaller.

Within a language, the difference between the structures of written and oral texts may vary from language to language, and for Russian, with its free word order, its simple system of tenses, its absence of articles and abundance of particles and interjections (ready-made fillers), the difference is smaller than for English. Typologically, the standard Russian language is oriented towards oral communication to a greater degree than the English language. Therefore the English oral text proves to be typologically closer to Russian (both oral and written) than to the English standard written text.

#### R E F E R E N C E S

Labov, W. The Study of Nonstandard English. Urbana, 1975.

Launsbury, F.G. Pausal, Juncture and Hesitation Phenomena.

- In: Psycholinguistics. Baltimore, 1954.

Земская Е.А. (ред.). Русская разговорная речь. М., 1973.

Николаева Т.М. Новое направление в изучении спонтанной речи (О так называемых речевых колебаниях). - "Вопросы языкознания", 1970, № 3.

#### ТИПОЛОГИЧЕСКОЕ ИЗУЧЕНИЕ АНГЛИЙСКОЙ УСТНОЙ РЕЧИ

Б. Гаспаров, В. Паперно

#### Р е з ю м е

В работе, построенной исключительно на материале магнитофонных записей живой речи, ставится задача описания английской устной речи как особой системы, имеющей свои собственные внутренние закономерности строения. Наиболее общим свойством устной речи, отличающим ее от письменной речи, признается нелинейность строения, состоящая, во-первых, в постоянном взаимодействии двух каналов передачи (вербальной последовательности и мелодики), и во-вторых, в особой орга-

низации вербальной последовательности — многочисленных разрывах синтаксических связей, паузах, обрывах и пересечении тематических линий, — при которой оказывается невозможным установление жестких структурных связей между элементами этой последовательности. В результате в устной речи на первый план выступает свободное ассоциативное наложение элементов текста друг на друга; из совокупности этих свободных, не регламентированных жестко синтаксическими отношениями ассоциативных связей формируется особый смысл устного сообщения, непохожий на смысл письменного текста и непереводимый полностью в последний. Смысл устной речи характеризуется отсутствием четкой регламентированности отношений между отдельными семантическими компонентами и в то же время открытостью для инкорпорирования все новых потенциальных ассоциативных связей. Данное положение иллюстрируется в работе анализом двух устных текстов, один из которых взят из публичной, и другой — из повседневной речи. Далее рассматриваются более подробно те формальные приемы, при помощи которых говорящие формируют описанные свойства устного текста. Такими приемами, наиболее важными для английской устной речи, является широкое использование особых перебивов (филлеров), разрывающих линейную последовательность; разбиение речи на краткие отрезки, между которыми отсутствует синтаксическая связь; особое употребление артикля, появление специального квази-артикля; упрощение системы времен и нарушение согласования времен. Данные особенности обнаруживают значи-

тельное сходство с приемами, описанными для русской устной речи. Можно констатировать, что структура русской и английской устной речи в целом имеет гораздо больше общих черт, чем структура этих языков в их письменной форме. Кроме того, в русском языке письменная речь, благодаря таким ее свойствам как свободный порядок слов, отсутствие артикля и согласования времен, в целом не так сильно противопоставлена устной речи, как это имеет место для письменной и устной формы английского языка. Таким образом, открывается перспектива типологического описания общих структурных свойств устной речи и того, как эти свойства соотносятся с особенностями синтаксической структуры различных языков.



## К ЭВОЛЮЦИИ МЕТАЯЗЫКА СОВРЕМЕННОЙ ЛИНГВИСТИКИ (препозиционный элемент ЛИНГВО-)

А.Д. Дуличенко

I. Современная лингвистика характеризуется охватом широкого круга проблем и системным подходом к их решению. Установление новых объектов исследования, связей и закономерностей, возникающих между такими объектами, требует, само собой разумеется, обращения к терминотворчеству, т.е. словообозначению выявленных объектов и характеризующих их отношений. Процесс терминотворчества особенно нагляден в работах, написанных в русле считающихся или считавшихся перспективными направлений и методов исследования, срав. известные разновидности современного структурализма, давшие целые словари собственной терминологии; срав. также весьма интенсивное возрождение интереса к ономастике, в которой составляющие ее пласты (единицы) получают отдельное "имя": топонимы и их разновидности оронимы, спелеонимы, гидронимы, дрионимы, пелагонимы, лимнонимы, потамонимы, ойконимы, урбонимы/полисо-нимы; космонимы/астронимы, космотопонимы/астротопонимы; этнонимы, антропонимы, теонимы, зоонимы, псевдонимы. Многие из этих "имен" появились недавно.

II. Чтение современной лингвистической литературы постоянно сталкивает нас с массой новых терминов, толкования которых мы не найдем в словарях лингвистических терминов; последние, как известно, не успевают фиксировать весь этот поток и включают лишь "устоявшуюся" терминологическую лексику. Мы не ставим задачей представить здесь обзор или провести анализ нефиксированной лингвистической терминологии. Мы хотим обратить внимание на заметно активизировавшийся в последнее десятилетие в метаязыке нашей лингвистики компонент ЛИНГВО-, давший к настоящему времени целый класс новых терминов и имеющий большие терминопотенции в будущем. Знаменательное слово латинского языка, войдя в метаязык лингвистики, перестало быть таковым и превратилось в префиксоид, с помощью которого постепенно и неуклонно вытесняются неудобные для терминологических целей, как правило, двух- или трехкомпонентные словосочетания.

П.1. В "Словаре лингвистических терминов" О.С. Ахмановой (1966) фиксируется лишь один термин и одна форма с элементом ЛИНГВО-: лингвостилистика и лингвостилистический (термин лингвистика - особый случай!). В некоторых зарубежных лексикографических изданиях по лингвистике рассматриваемый тип терминов представлен несколько шире, как, например, в "Энциклопедическом словаре лингвистических терминов" Р. Симеона (Siméon, 1969, 773), где, помимо указанного, встречаем:

lingvoalveolarni glasovi	- лингвоальвеолярные звуки
lingvodentalni glasovi	- лингводентальные звуки
lingvofilozofija	- лингвофилософия
lingvogram	- лингвограмма

И эти термины не отражают, конечно, всей полноты "лингво"-терминологии.

П.2. Составленный нами список (далеко не полный и не окончательный) интересующих нас новообразований в функционально-семантическом плане неоднороден. Наиболее продуктивна и многочисленна по количеству составляющих ее единиц группа, в которой элемент ЛИНГВО- присоединяется к терминам, отражающим конкретные дисциплины гуманитарного (преимущественно) и естественного циклов:

лингвогеография  
лингвоисториология  
лингвопалеонтология  
лингвосемиотика  
лингвосоциология  
лингвостатистика  
лингвостилистика  
лингвострановедение  
лингвотекстология  
лингвотипология  
лингвофилософия.

Элемент ЛИНГВО- может присоединяться также к словам, выражающим различные объекты единичного или собирательного характера или действия, связанные с ними, как в случаях: лингвопроект и лингвопроектирование, лингвосоциум, лингвокод и лингвокодирование, лингвограмма и др. Экспансии элемента ЛИНГВО- подверглись также композиты типа национальная общность, исторический анализ, срав.: лингво-национальная общность, лингвоисторический анализ [при отсутствии (пока!) лин-

гвоистория] и даже "лингвоархеологическая" (пока в кавычках!) реконструкция.

Ш. Чье же место занял рассматриваемый нами элемент?  
Нетрудно заметить, что ЛИНГВО- постепенно вытесняет:

1) компонент лингвистический в составе терминологических композит -

лингвистическая география > лингвогеография

лингвистическая историология > лингвоисториология

лингвистическая палеонтология > лингвопалеонтология;

2) компонент языковой/языковый в составе терминологических композит -

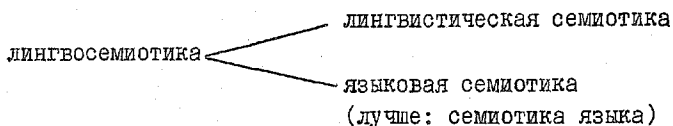
языковой проект > лингвопроект

языковой социум > лингвосоциум.

Срав. также: языково-национальная общность > лингво-национальная общность и под.

Ш.І. Компоненты лингвистический, т.е. относящийся к лингвистике, науке, и языковой, т.е. относящийся к языку как объекту лингвистики, несмотря на очевидную разность их функциональной отнесенности, в современной лингвистике практически не дифференцируются. Достаточно обратиться к словарю О.С. Ахмановой (1966), чтобы убедиться, что между терминами лингвистическая география, т.е. узкой лингвистической дисциплиной, разделом, и языковая география, т.е. фактом территориального распространения специфических явлений конкретного языка или диалекта, стоит знак равенства; срав. также расшифровку термина лингвосоциум как "лингвистический коллектив" вместо "языковой коллектив", "коллектив данного языка" (ВЯ, 1977, 6, с. 137). Более того, компоненты лингвистический и языковой синонимируют не только между собой, но и с компонентом национальный, срав. часто встречающиеся в социолингвистике и в этнографии сочетания: языковое меньшинство, лингвистическое меньшинство и национальное меньшинство!

Видимо, различие, о котором мы говорим, не столь существенно (срав. также в этнографии обычную подмену этнографическая группа вместо этническая группа!) и носит скорее формальный характер. Новый элемент ЛИНГВО- как бы синтезирует в себе два аспекта рассматриваемого вопроса, в результате чего возможны корреляции типа:



Такое соотношение характерно для первой группы "лингво"-терминов (лингвогеография ... лингвофилософия), в которой второй словообразовательный компонент при самостоятельном его употреблении выступает в значении "дисциплина", "раздел", "отрасль", "наука". Что касается терминов типа лингвосоциум, лингвопроект, то своим коррелятом они имеют, как правило, языковой (социум, проект). Однако и здесь нет прочного соответствия, срав. возможности трактовки термина лингвокод как 1) языковой код и 2) лингвистический код и под.

Таково состояние конкуренции по линии ЛИНГВО- - ЛИНГИСТИЧЕСКИЙ - ЯЗЫКОВОЙ на сегодняшний день.

III.2. Необходимо отметить тот факт, что роль конкурентов элементу ЛИНГВО- могли бы составить знаменательные слова - язык (славянского происхождения) и глотта/глосса (греческого происхождения), как и ЛИНГВО-, способные становиться префиксоидами в составе сложного терминологического целого. На раннем этапе развития отечественной науки о языке (конец XVIII - начало XIX вв.) для ее обозначения появляется калькированный термин языковедение (позднее стали употреблять и производные: языковедческий, языковед); препозиционный элемент ЯЗЫКО- получает в дальнейшем некоторое распространение: появляется дублет языкознание; в лингвистической литературе XIX в. можно было встретить и термины языкоиспытатель - в значении деятеля, языкотворчество (и производное языкотворческий) - в значении процесса. Конкуренцию препозиционному элементу ЯЗЫКО- в XIX в. составил греческий элемент ГЛОТТО-/ГЛОСС-. Так, конкуренция терминов глоттология и языковедение/языкознание решена в пользу последнего (последних). Зато сохранились термины глоттохронология, глоттогенез (реже - глоттогония, чаще - глоттогонический); срав. также: глоссология, глоссарий и в настоящее время - глоссематика.

Термины с элементом ЛИНГВО- появляются значительно позже, когда образования на ЯЗЫКО- и ГЛОТТО-/ГЛОСС- прочно вошли в метаязык лингвистики; этим объясняется излишность введения, например, для терминологической семемы "наука о языке" терминов типа лингвоведение (срав. литературоведение!)

или лингвология; к тому же для этих целей послужило и вскоре заимствованное лингвистика, конкурирующее до сих пор с языковедение и языкознание.

IV. Строение гипотетического лингвоведения, которого мы только что коснулись, и реального лингвострановедения и под. не должно вызывать недоумения, поскольку подобная гетерогенность (сочетание "чужой элемент + свой элемент") терминологических значений в метаязыке лингвистики (как, впрочем, и других наук) имеет довольно длительную традицию и широкое распространение, срав.: субполе, метаязык, метаописание, параязык, противоязык и др.; срав. также другой компонент филологии — литературоведение. Из других областей: металловедение, машиноведение, планетоведение, музыковедение, библиотековедение и т.д.; автодело, автосообщение; радиовыставка, радиосеть и др. При обратном соотношении ("свое + чужое") компонентов: словоформа, словоэлемент, звукотехника и под.

У. Что касается причин экспансии элемента ЛИНГВО- в метаязыке современной лингвистики, то их, как нам кажется, несколько:

1) постоянно проявляющая себя в сфере терминологии тенденция к экономии и удобству выражающих средств, срав. вместо происхождение языка — глоттогенез, вместо лингвистическая стилистика или стилистика языка — лингвостилистика; вместо статистический анализ языка — лингвостатистический анализ или лингвостатистика; вместо текстологическое изучение языка — лингвотекстологическое изучение или лингвотекстология и т.д.;

2) постоянно проявляющаяся необходимость в унификации аналогичных терминорядов; имеющие отношение к языку (или лингвистике) методы, разделы и т.д., как, например, статистика, типология, семиотика и под. маркируются, дабы реально выразить это отношение, элементом ЛИНГВО-;

3) не в последнюю очередь на интенсификацию использования элемента ЛИНГВО- в метаязыке нашей лингвистики повлияла аналогия; в ряде наук терминоряды маркируются элементом, являющимся главным компонентом обозначения таких наук, срав. в этнографии ЭТНО-: этноним/ика/ия, этноботаника, этнопсихология, этногеография, этнотопоним/ия, этнодифференцирующий (признак), этносоциальные (организмы, процессы, образования), этноэволюционные (процессы), этнокультурная (информация), -ое

(единство, различие, явление), -ий (контакт, процесс), этно-расовые (смещения), этноизолирующий (фактор) и т.д.; срав. также ПСИХО- в психологии: психогигиена, психопатология, психотерапия, психофизиология, психоневрологический, психо-профилактический и др.

Примеров можно было бы привести великое множество. Пре-позиционные элементы ЭТНО-, ПСИХО-, ХИМИО-, БИО- и под. маркируют термины, относя их к наукам, обозначения которых начинаются с соответствующих элементов.

VI. Интересно отметить, что элемент ЛИНГВО- активен преимущественно в препозиции. Чем дальше от начала слова (термина), тем все более ЛИНГВО- в форме ЛИНГВ- начинает конкурировать с элементом ЯЗЫК-/ЯЗЫЧ-, срав.:

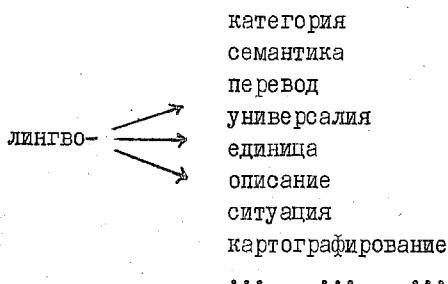
билингвизм	- двуязычие
монологизм	- одноязычие
мультилингвизм	- } многоязычие
полилингвизм	

В случаях, когда семема, выражающаяся элементом ЛИНГВО-, занимает постпозицию в термине, интересующий нас элемент во все невозможен, срав.: хиндиязычный, франкоязычный; написанно-языковое (строительство), культурно-языковое (планирование). Причина кроется, видимо, в невозможности адекватной формы от ЛИНГВО-, срав., однако этнический < ЭТНО-.

VII. Сказанное дает возможность сделать некоторые прогнозы относительно судьбы интересующей нас группы терминов и дальнейшего распространения препозиционного элемента ЛИНГВО-.

I) Нет сомнения в том, что большинство отмеченных в данной статье "лингво"-терминов (лингвосемиотика, лингвострановедение, лингвопалеонтология ...; лингвопроект, лингвокод и под.) прочно войдет в метаязык современной лингвистики и как результат этого окажется в будущих словарях лингвистических терминов. Эту мысль подкрепляет и тот факт, что часть таких терминов практически уже общепринята, срав., например, использование их в названиях научных статей ("Некоторые вопросы соотношения лингво- и этнографических исследований", "Роль санскритских элементов в лингвоисторическом анализе"), разделов монографий и даже сборников ("Лингвотипологические исследования", М., 1973; "Лингвострановедческий аспект преподавания русского языка иностранцам", М., 1974 и др.).

2) Элемент ЛИНГВО- проявляет все большую активность к распространению; тем самым можно ожидать в недалеком будущем появления ряда новых "лингво"-терминов, как, например, в следующем случае:



Срав. также предложенные несколько ранее автором настоящей статьи (в целях уточнения и совершенствования метаязыка лингвистики) термины лингвосистема и лингвоструктура вместо неудобных и нетерминологичных система языка/языковая система и соответственно структура языка/языковая структура.

В связи с разрабатываемым нами вопросом о названиях языков и их разновидностей как особом пласте терминов встала необходимость введения ряда "лингво"-терминов типа: лингвоним "название языка" (и производные: лингвонимия "совокупность лингвонимов", лингвонимика "метаязыковый раздел лингвистики, занимающийся изучением лингвонимии"); вместо тяжелых и противоречащих принципу терминологичности предложены были также термины типа: лингвогруппа (срав.: языковая группа, группа языков), лингвосемья (срав.: языковая семья, семья языков) и т.д.; срав. в этой связи также: славянская лингвогруппа вместо славянская группа языков, славянская языковая группа; индоевропейская лингвосемья вместо индоевропейская языковая семья, индоевропейская семья языков и под. (Дуличенко, 1978; Duličenko, 1973). Термин лингвоним, например, начинает входить в научное употребление, см. (Языковая ситуация, 1975).

ПРИМЕЧАНИЕ: Определение современный в сочетании с лингвистика, неоднократно появляющееся в статье, указывает лишь на хронологический признак: современная лингвистика - лингвистика наших дней.

## Л и т е р а т у р а

Ахманова О.С. Словарь лингвистических терминов. М., 1966.

Дуличенко А.Д. Очерки по общей и русской лингвонимике.

1. К метаязыку лингвистики: лингвонимы как особый класс терминов. - Лингвистическая семантика и семиотика. Вып. I. Тарту, 1978.

Языковая ситуация в странах Африки. М., "Наука", 1975.

Duličenko A.D. La lingvonimiko (ĝia esenco kaj problemoj).

- Scientia Revuo, Beograd, 1973, vol. 24, N 2-3.

Simeon R. Enciklopedijski rječnik lingvističkih naziva. I-II. Zagreb, 1969.

### ON THE EVOLUTION OF THE METALANGUAGE OF PRESENT-DAY LINGUISTICS

(the Prepositive Element LINGVO-)

A. D. Duličenko

#### S u m m a r y

Within the evolving systems of linguistic terms the element LINGVO- is conspicuous for its activity. It has already given rise to a whole class of new terms and possesses a great term-producing potentiality for the future: lingvo-geografija, lingvosemiotika (names of concrete disciplines, methods), lingvoprojekt, lingvokod (names of things). LINGVO- has ousted the components lingvističeskij and jazykovoj in compounds of the type lingvističeskaja semiotika - lingvosemiotika, jazykovoj projekt - lingvoprojekt. It is only in the case of the terminological sememe "the science of language" that terms of the type lingvovedenje and lingvologija are superfluous because already in the 19th century the Russian JAZYKO- (jazykoznanije, jazykovedenje) and the Greek GLOTTO- (glottologija) came to be used for this purpose, as well as the later borrowing lingvistika. The spread of the element LINGVO- has been encouraged by a tendency towards economy and convenience (lingvotekstologija instead of tekstologičeskoe izučenje jazyka), the need for unification (sets of terms pertaining to language are marked by the element LINGVO-), analogy with other sciences (cf. the sets of terms involving ETNO-, PSIKHO-, etc.). It may be assumed that a large number of terms containing the element LINGVO- will appear in the future: lingvokategorija, lingvosemantika, lingvopisanije, etc.; cf. likewise the terms introduced by the present writer: lingvonim "designation (name) of a language" and its derivatives, also lingvogruppa, lingvosem'ja, etc.



# PROBLEMS OF RENDERING THE SEMANTIC STRUCTURE OF THE UTTERANCE IN TRANSLATION

Urve Lehtsalu, Gustav Liiv

In translating from one language into another the translator is faced with the task of rendering the semantic structure of the sentence, i.e. functional sentence perspective.

The theoretical foundations of functional sentence perspective were laid down by the Czech linguist V. Mathesius (1882-1945) (Матезиус, 1967). Rejecting the psychological interpretation of the semantic structure, V. Mathesius introduced the term "functional sentence perspective" into linguistics. The linguistic theory of the semantic structure was further developed by members of the Prague school of linguists, such as J. Firbas, E. Dvořáková, and others (Firbas, 1965; Dvořáková, 1972).

The problem of functional sentence perspective has been dealt with in several works by Soviet authors (Ильин, 1961; Распопов, 1961; Колманский, 1965). Attempts have been made to analyse the relation between the semantic structure of the utterance and the grammatical structure of the sentence (Гак, 1968). In recent years the problem of the semantic structure of the utterance has aroused great interest also in the theory of translation (Черняховская, 1976; Левякова, 1976).

Most utterances consist of two sections, one of them containing the starting point of the statement (i.e. the theme) and the other the new information (i.e. the rheme).

The essential feature of the theme is the lowest degree of communication dynamics (CD) (Firbas, 1964). Depending on the context it may convey information already known or something not mentioned before (Ильин, 1965, 203).

On the other hand, the rheme is characterized by the highest degree of communication dynamics. It may include either new information or something already known to the reader or the listener but serving as the centre of the pre-

dication ( Шенякова, 1976, 59).

The means of expressing a thematic or a rhematic quality of a word or phrase in a sentence depend to a great extent on the grammatical structure of the given language and must differ considerably according to that structure (Ильин, 1965, 197). Therefore the relation between the syntactic structure of the sentence and functional sentence perspective has been receiving ever growing attention in the theory of translation.

Differences in the grammatical and lexical means of expressing a thematic or rhematic quality are especially marked in languages of different typology such as English with its analytical structure and Estonian, which belongs to synthetic languages.

As in other synthetic languages, in Estonian its relatively free word order can be used to mark the rheme and the theme of the sentence. If we compare the sentences 'Raamat on laual' and 'Laual on raamat', we perceive that in the former sentence the highest amount of information is expressed by the adverbial modifier 'laual', whereas in the latter sentence the subject 'raamat' conveys the new information. Word order serves as a most adequate means of pointing out the rheme in the Estonian sentence. Thus, the rhematic quality of the grammatical subject can be indicated by its position towards the end of the sentence, e.g.:

Samal õhtupoolikul ja umbes samal ajal võitlesid klassitoas kaks koolipoissi. (O. Luts)

Due to its more rigid word order the rising degree of communication dynamics is less characteristic of English. English word order is governed by grammatical principles. Therefore other grammatical and lexical means are used in English to express the semantic structure of the utterance. At the same time it has been pointed out that there is a tendency also in English to construct utterances with a growing degree of communication dynamics (Черняховская, 1976, 28). As the grammatical subject in English usually stands at the beginning of the sentence, it generally serves to express the theme (Mathesius, 1964, 61). Thus the semantic and the grammatical structures of the sentence more often coincide in English than in Estonian. In translation from English into Estonian these differences have to be compen-

sated by syntactic changes and lexical substitutions.

It is the aim of the present paper to provide a brief survey of the principal means of rendering the semantic structure of an English utterance in translation.

In Present-day English there are several ways of expressing the highest degree of communication dynamics, i.e. the rheme of the sentence.

The rheme in an English sentence often follows the construction 'there is'. In the Estonian translation the rheme will be transferred to the end of the utterance, the adverbial modifier being placed at the beginning, e.g.:

There were never many girls at all the football games. (J.D. Salinger)

Jalgpallivõistlustel ei olnud kunagi palju tüdrukuid.

There is a soul in each of us. (O.Wilde)  
Meis igaühes on hing.

There was something celestial about it... (J.London)

Tolles hääles oli midagi taevalikku...

This change in the grammatical structure of the sentence often entails concretization of the predicate verb, e.g.:

There was a rustle of chirruping sparrows in the green lacquer leaves in the ivy. (O.Wilde)

Luuderohu läikivate lehtede rohelist kostis varblaste siutsumist ja sabinat.

There was a stream alongside the road. (E.Hemingway)

Tee kõrval jooksis jõgi.

Another analytical method of indicating the rheme in English, finding its parallel in French, is the construction 'it is... that', 'it is... who' and 'it is... which' with a word or phrase representing the rheme placed between the words 'it is' and the word 'that' ('who', 'which'). This construction can indicate the rhematic quality of any part of the sentence except the predicate. In the Estonian transla-

tion the rheme can be placed towards the end of the sentence, e.g.:

It was of himself and his own future that he had to think. (O.Wilde)

Tal tuleb mõelda iseendale ja oma tulevikule.

It was the missionary who did the image breaking. (J.London)

Pilte hävitasid misjonärid...

Depending on the emotional charge a similar construction can be found in Estonian, though it is less characteristic of the Estonian language, e.g.:

It was his beauty that had ruined him.  
(O.Wilde)

See oli ilu, mis oli ta hävitanud.

A means of indicating the rheme of a sentence by the grammatical subject may be the indefinite article (the zero article in the plural). Owing to its basic meaning of indefiniteness, the indefinite article serves to signalize the new information in the sentence, i.e. the rheme. As in Estonian the highest degree of communication dynamics is placed towards the end of the utterance, changes in word order have to be made in translation, e.g.:

A horse moved and neighed by a bush.

(A.Sillitoe)

Ühe põõsa juures liigatas ja hirnatas hobune.

A grasshopper began to chirrup by the wall.

(O.Wilde)

Müüri ääres hakkas siristama rohutirts.

In the grass, white daisies were tremulous.

(O.Wilde)

Rohus värelesid valged karikakrad.

In the plural a partial subject often occurs in Estonian, e.g.:

Primroses grew to each side of the track.

(A.Sillitoe)

Kummalgi pool teed kasvas nurmenukke.

There were worse cripples around. (S.Bellow)

Leidus inimesi, kes olid hoopis hullemini haiged.

Sometimes it would be impossible to place the highest degree of communication dynamics at the end of the utterance, and the rhematic quality of the subject in the Estonian language has to be indicated by some lexical means (words like 'keegi', 'mingi', 'üks'), e.g.:

A goat hopped up on one of the carts and then to the roof of the diligence. (E.Hemingway)

Üks kits hüppas kaarikule ja sealt postitõlla katusele.

A change, like a shift of wind, overcame the Judge. (T.Capote)

Mingi muutus toimus kohtunikus, nagu oleks tuul teisale pööranud.

The rheme in an English utterance can also be indicated by words like 'even', 'only', 'merely', 'just', 'namely', 'such as', 'rather than', 'at least', etc. Besides having a semantic function of its own, such a word or phrase serves to point out the highest degree of communication dynamics, e.g.:

His cool, white, flowerlike hands, even, had a curious charm. (O.Wilde)

Isegi ta külmadel, valgetel lillõrnadel kätetel oli imelikku võlu.

As similar means of indicating the rheme are characteristic also of Estonian, they do not present difficulties from the point of view of translation.

The rheme of the English utterance, expressed by the grammatical subject of the sentence, can sometimes be indicated by inversion. As in Estonian the end-position of the subject is as common as its initial position (Valgma, Remmel, 1968, 252) the retention of the semantic structure of the English sentence does not present difficulties to the translator. At the same time the Estonian utterance may sound less emphatic, e.g.:

Lying on the floor was a dead man, in evening dress, with a knife in his heart. (O.Wilde)

Põrandal lebas õhturiietuses laip, nuga südames.

Coming at a fair rate of speed into the square was an open car. (I.Shaw)

Väljakule sõitis küllaltki suure kiirusega lahtine auto.

The highest degree of communication dynamics can also be expressed by the prepositional object in a passive construction. In the Estonian translation the rheme is usually expressed by the subject of the sentence which is placed after the object or adverbial modifier and the predicate, e.g.:

His night had been untroubled by any images of pleasure or pain. (O.Wilde)

Tema ööd polnud seganud ükski lõbude ega piinade viirastus.

At the same time numerous cases can be found where the rheme is expressed by the subject of the passive sentence, e.g.:

Two globe-shaped china dishes were brought in by a page. (O.Wilde)

Teenriipoiss tõi kaks kerakujulist portselanumat.

In the Estonian translation the highest degree of communication dynamics is expressed by the object in the end-position.

In English there are also several ways of referring a word or phrase to the lowest degree of communication dynamics, i.e. the theme.

As already mentioned above, the theme of an English utterance is often expressed by the grammatical subject of the sentence, particularly in passive constructions. In the Estonian translation the passive voice is mostly replaced by the personal mood, the theme being expressed by the object of the Estonian sentence, e.g.:

They were caught up from the warm troughs of sleep by Seaton's rough voice at seven o'clock. (A.Sillitoe)

Kell seitse hommikul raputas Seaton'i rääme häälel nad soojust unehõlmast lahti.

The theme of an English utterance is frequently ex-

pressed by a grammatical subject denoting a lifeless object but attached to a predicate verb which is generally used with nouns denoting living beings. In the Estonian translation the theme is expressed by an adverbial modifier, e.g.:

Classes were to begin within two days and the past week had been a steady stream of new arrivals. (A.Wilson)

Kahe päeva pärast pidi õppetöö algama ja seetõttu oli uustulnukaid terve möödunud nädal ühtejärke kokku voolanud.

There is another means of indicating the theme which includes both grammatical and lexical phenomena. This is a parenthesis introduced by the prepositional phrase 'as for' or 'as to' while in the main body of the sentence there is a personal pronoun referring to the noun in the prepositional phrase, e.g.:

As for Alan Campbell, his suicide had been his own act. (O.Wilde)

Mis Alan Campbelli surmasse puutub, siis oli see tema enda tegu.

As for believing things, I can believe anything, provided that it is quite incredible. (O.Wilde)

Mis puutub uskumisse, siis suudan uskuda kõike eeldusel, et see on täiesti uskumatu.

A word or phrase representing the theme may also be placed at the beginning of the sentence as a loose part of it announcing what the sentence is going to be (Ильин, 1965, 202). In the body of the sentence it is repeated by a demonstrative pronoun, less frequently by the same noun. In the Estonian translation the same construction can be used, with certain lexical emphatic means sometimes added, e.g.:

A new life! That was what he wanted. (O.Wilde)

Uus elu! Just seda oli tal vaja.

One's own soul and the passions of one's friends - those were the fascinating things in life. (O.Wilde)

Su enda hing ja su sõbra kired - need on kõige võluvamad asjad siin elus.

The gypsy, he thought. He is truly worthless.  
(E.Hemingway)

Mustlane, mõtles ta. Mustlane on lontrus.

The definite article may also serve as a marker of the-  
matic quality, e.g.:

The flies buzzed round the table and crawled  
over the stained cloth. (O.Wilde)

Kärbsed sumisesid laua ümber ja ronisid mää-  
dunud linal.

In the Estonian translation the front-position of the  
subject marks the theme.

However, the grammatical subject of the English sen-  
tence often lacks a formal indicator of its rhematic or the-  
matic quality. The semantic structure must therefore be con-  
cluded from the context, e.g.:

Music had stirred him like that. Music had  
troubled him many times. (O.Wilde)

In the first sentence the comparison 'like that' has a the-  
matic quality as it links the sentence to what has been said  
before, the word 'music' conveying the new information. In  
the second sentence 'music' is already the starting point of  
the utterance, i.e. the theme. In the Estonian translation  
this is reflected by the word order:

Samasugusel viisil oli teda varem liigutanud  
muusika. Muusika oli teda sageli erutanud.

This leads to another interesting problem concerning  
functional sentence perspective, called theme dynamics, i.e.  
the patterns by which the themes reoccur in a text and by way  
they run through the successive sentences of a text (Daneš,  
1970). Problems of retaining theme dynamics in translation  
have already attracted the attention of linguists working in  
the field of the theory of translation (Шевякова, 1976, 56-  
65). Further contrastive studies in this field will also pre-  
sent interest from the point of view of translation from  
English into Estonian.



## REFERENCES

- Daneš, F. One Instance of Prague School Methodology. Functional Analysis of Utterance and Text. Method and Theory of Linguistics. The Hague, 1970.
- Dvořáková, E. On the English and Czech Situational Adverbs in FSP. - Brno Studies of English, vol. 4, 1972.
- Firbas, J. On Defining the Theme in Functional Sentence Analysis. - Travaux Linguistiques de Prague, vol. 1, 1964.
- Firbas, J. A Note on Transition Proper in Functional Sentence Analysis. - Phil. Prag., 8 (47), Nr. 2-3, 1965.
- Mathesius, V. On Linguistic Characterology with Illustrations from Modern English. - Prague School Reader in Linguistics. Bloomington, 1964.
- Valgma, J., Rimmel, N. Eesti keele grammatika. Tallinn, 1968.
- Гак В.Г. Проблемы лексико-грамматической организации предложения. Автореферат диссертации на соискание ученого степени доктора филологических наук. М., 1968.
- Ильин Б.А. Развитие способа выражения смыслового предиката в английском языке. - Вопросы германского языкознания. М., 1961.
- Ильин Б.А. Строй современного английского языка. М., 1965.
- Колшанский Г.В. Логика и структура языка. М., 1965.
- Матезюс В. О так называемом актуальном членении предложения. Основные функции порядка слов в чешском языке. - Пражский лингвистический кружок. М., 1967.
- Распопов И.П. Актуальное членение предложения. Уфа, 1961.
- Черняховская Л.А. Перевод и смысловая структура. М., 1976.
- Шевякова В.Е. Актуальное членение предложения. М., 1976.

ПРОБЛЕМЫ ВЫРАЖЕНИЯ СМЫСЛОВОЙ СТРУКТУРЫ  
ПРЕДЛОЖЕНИЯ В ПЕРЕВОДЕ

У. Лехтсалу, Г. Лийв

Р е з ю м е

В статье рассматриваются типичные способы выражения логического подлежащего и логического сказуемого в английском предложении. При сопоставлении семантической структуры простого предложения в английском и эстонском языках указывается на главные различия в ее синтаксическом выражении. Авторы останавливаются подробнее на способах передачи темы и ремы английского высказывания в эстонском языке.

К ВОПРОСУ ОБ ОЦЕНОЧНОСТИ ОБЩЕСТВЕННО-  
ПОЛИТИЧЕСКОЙ ЛЕКСИКИ  
(На материале передовиц "Dagens Nyheter")

И. Мулламаа

Выражение оценки в передовицах тесно связано со своеобразием использования языка в этом жанре и в функциональном стиле публицистики в целом. Предметом и сферой функционирования публицистического стиля речи и газетного подстиля является общественная жизнь и ее разные аспекты — экономический, политический, идеологический, культурный, спортивный и др. В жанре передовиц находят отражение все названные аспекты, пропущенные через призму отношения к ним редакции и авторов передовых статей. Главной целью передовиц является не просто информация, а комментарии, в частности оценки со стороны авторов передовиц, а шире — редакций, партий или кругов, которые стоят за газетой.

В данной статье анализируются некоторые аспекты оценочности общественно-политической лексики в передовицах самой крупной утренней газеты Швеции "Dagens Nyheter" за 1976 год.

Поскольку в рассматриваемых передовицах отражаются самые разные темы общественной жизни, лексика в них неоднородна. Специфика газеты — в ее массовости, в отсутствии прямой, контактной связи между коммуникантами, и это требует доступности, общепонятности языка, определенных ограничений словоупотребления как в газете в целом, так и в передовицах. В то же время передовицы характеризуются своеобразной приподнятостью, так как все, что там имеется, рассматривается как авторитетный, исполненный большого общественного значения материал, который должен получить соответствующую речевую форму.

В исследуемых передовицах преобладала общественно-политическая лексика, хотя тематика в них, как было отмечено, является весьма разнообразной. В эту лексику входят названия явлений и понятий из области международной, внутривнутриполитической, общественной и экономической жизни.

Семантика общественно-политической лексики характеризуется абстрактностью, которая связана с преобладанием в жанре передовиц понятийно-логической формы мышления. В зависи-

мости от коммуникативного задания передовиц (воздействие, комментирование, оценка), эта интеллектуализированная лексика получает в них чаще всего оценочные, а также экспрессивные и эмоциональные функции (ср. Зубов, 1973:63). Как представляется, главной функцией общественно-политической лексики в передовицах является функция оценки, а экспрессивность и эмоциональность исследуемых здесь лексических единиц нужно считать средствами усиления выразительности оценки, а тем самым усиления ее воздействия.

Использование общественно-политической лексики в средствах массовой информации и пропаганды находится в прямой зависимости от идеологии и политических взглядов партий и кругов, стоящих за этими средствами, а также от взглядов отдельных авторов.

В общественно-политической лексике выражается как влияние идеологии на функционирование языка, так и, как результат этого влияния, — отражение идеологии в семантике языковых единиц. Это значит, что общественно-политическую лексику и ее использование в речи можно считать идеологизированными. Идеологизированной является прежде всего та часть общественно-политической лексики, которая имеет оценочное значение или коннотации. Такое содержание может быть присуще новым словам, образованным в основном при помощи аффиксов и словосложения (Напр. *medbestämmande*, *medbestämmanderätt*, *medbestämmandepolitik*, *löntagarfonder*). Эти слова возникают в связи с изменениями, происходящими в общественной жизни. Вместе с тем, идеологически окрашенное значение приобретают и слова, прочно вошедшие в общественно-политическую лексику (ср. Крючкова, 1976:14-18).

Для идеологизированной лексики характерно, что она в смысловом или оценочном планах по-разному воспринимается представителями разных идеологий.

Возможны, однако, случаи, когда слова имеют постоянную оценочную коннотацию, характер которой одинаково воспринимается большинством читателей газеты. Читатель в таких случаях реагирует прежде всего на общественно фиксированное созначение слова и тем самым даже незаметно для себя принимает точку зрения автора и на те денотаты (факты и события), которые данным словом называются. К словам с фиксированным оценочным созначением относятся, например *demokrati*, *demokratisering*, *färdhet*, *miljö*, *miljöhänsyn*, *miljöpolitik*, *sysselsättning*,

sysselsättningspolitik, trygghet, trygghetsland, trygghetsreform, välfärd, välfärdslandet, aggression, ockupation.

Это т.н. слова-сигналы, в которые в зависимости от целей автора высказывания можно каждый раз вкладывать новое содержание (Костомаров, 1967:59-60). Часто использование таких слов в передовицах связано с тем, что их прагматическая (оценочная) информация устойчива (Klaus 1971:56 ff.). В этих словах предметнологическое (дескриптивное) значение почти стерто, и они используются в основном как знаки (ярлыки) положительной или отрицательной оценки. С их помощью газеты производят т.н. *begging the question*<sup>22</sup> т.е. навязывание такой оценки общественных и коммерческих явлений, которую следовало бы доказать соответствующей аргументацией (Fredriksson, 1969:45-63). Например, такие слова и выражения, как *demokrati*, *demokratiskt samhälle*, *de demokratiska fria och rättigheterna*, *välfärdsamhälle*, *den fria världen*, *fri företagsamhet* в передовицах в основном имели чисто сигнальную функцию, передавая положительное или отрицательное отношения автора к явлениям, которые этими словами обозначаются, но содержащаяся в этих словах оценка оставалась, как правило, без экспликации. Например, в предложении "När har vi en hyggligt fungerande demokrati, där nya socialiseringssteg kan och skall diskuteras" /DN 15/7 /, где характеризуется обстановка в Швеции, остается без доказательств факт наличия демократии и ее положительности. Эти моменты не ясны и из контекста. Значит, автор статьи навязывает свою точку зрения, используя положительные коннотации слова *demokrati*.

Вместе с тем, возможны случаи, когда слова имеют нефиксированное оценочное созначение, и у представителей разных идеологий название явлений и фактов такими словами связано с противоположными оценками как денотатов, так и самих понятий. К таким словам относятся, например, *socialism*, *kommunism*, *kapitalism*, *den fria världen*, *stormakt*, *socialdemokrat*, *kommunist*, *vänsterkrafter*, *högerkrafter*, *vänsterkoalition*, *högerkoalition*.

Конечно, нельзя утверждать, что вся общественно-политическая лексика имеет оценочные коннотации. Так, большинство

<sup>22</sup> to beg the question - to take for granted the matter in dispute (Standard Dictionary of the English Language, 1963:127).

терминов экономики лишены оценочности. Напр.: *bruttonationalprodukt, inkomstskatt, indexreglering, kollektivavtal, marginalskatt*. Эти слова включены в терминологическую систему соответствующей науки. Оценочность характерна, в основном, для такого раздела общественно-политической лексики, который не может быть терминологией в силу неоднозначности лексем для разных групп общества. И эта неоднозначность выявляется именно в оценочном плане, а ориентированность, характер и основание оценки определяются широким экстралингвистическим контекстом. Такие слова, как напр. *företagademokrati, samarbetspolitik, närdemokrati, välfärdssamhälle* в передовицах понятийно не определены. Но в них содержится созданная прессой оценочная коннотация.

Использование ряда слов общественно-политической лексики в исследованных передовицах свидетельствует о тенденции к выражению этими словами замаскированной оценки или модификации оценки. Они представляют собой замену некоторых классических терминов, которые обладали определенной оценочностью. Часто это новые слова, в которых идеологическая острота стирается. Так, например, почти все без исключения авторы передовиц использовали слова *arbetstagare, löntagare, de anställda* вместо *arbetare* и *arbetsklass*; слова *arbetsgivare* и *kapitalägare* вместо *kapitalist*; слова *recession, nedgång, konjunktursvacka* вместо *kris*.

Такая манипуляция - замена исходных слов с неприемлемой для авторов коннотацией словами с приемлемой коннотацией, позволяет считать эти новые слова оценочными. Лингвисты называют это мелiorацией понятий (Шувалов, 1975:141).

При подобной замене объект высказывания, денотат, остается прежним, но вследствие сужения значения (ср., напр. *arbetare - löntagare, kapitalist - arbetsgivare* достигается эффект "снятия" сильной нежелательной коннотации или идеологической напряженности слова.

Для выражения оценочного отношения в общественно-политической лексике шведского языка очень широко используют сложные слова. В основном комбинируются слова, которые уже вошли в общественно-политическую лексику, с оценочными словами (они могут быть как определяющими, так и определяемыми). Часто эти оценочные слова заимствуются из разговорной речи, например, *fusk* в *skattefusk* и *valfusk*; *pryl* в *prylsamhälle, prylkonsumtion*. В результате появляются новые понятия и

новые оценочноокрашенные слова, которые экспрессивны и эффективны как средства воздействия.<sup>ж</sup>

К настоящему времени этот способ расширения фонда оценочной лексики является наиболее продуктивным в шведском языке. Об этом свидетельствуют многочисленные примеры из газетных статей, в частности, передовиц "Dagens Nyheter".

Напр.: konsumtionssamhälle, medansvar, medbestämmandelag, paradargument, prylsamhälle, samförståndspolitik, skrämselkampanj, skrämselpropaganda, trygghetslag, trygghetslagstiftning, välfärdspolitik, välfärdssamhälle.

Интересную экспрессивную разновидность использования сложных слов для выражения оценки представляют случаи, когда они употребляются в переносном смысле, то есть, когда на передний план выступает не столько название предмета, сколько его характеристика, оценка (ср. Гуськова, 1977:2 и след.). Несвязанность переносов непосредственно с темой статьи получает свое максимальное выражение именно в оценочных словах общественно-политического характера.

Напр. brobyggare, gräsetanivå, gummiormel, högerspråke, jargonsklackspråk, kautschuckparagraf, vänsterspråke.

С появлением на полосе газеты эти слова производят особенно сильное экспрессивное воздействие на читателя, но при длительном использовании они теряют экспрессивность, тем не менее сохраняя оценочные коннотации.

Выражение оценки с помощью общественно-политической лексики можно охарактеризовать как тенденцию ассоциировать слова-ярлыки с соответствующими денотатами, к которым авторы имеют положительное или отрицательное отношение. В таких словах предметно-логическое значение отступает на второй план, и на первый план выходит оценочная коннотация.

---

<sup>ж</sup> Ср. использование сложных слов в рекламе (Kleivane, 1974: 62).

## Литература

- Гуськова Е.В. Вторичная контекстуально обусловленная номинация в художественном тексте. Автореф. канд. дисс., М., 1977.
- Зубов А.В. Относительно некоторых лингвистических средств оценки в прессе (на материале английского языка). — Сборник научных трудов МГПИИЯ. Вып. 72. II., 1973.
- Костомаров В.Г. Слова-сигналы. — "Русская речь", 1967, № 3.
- Крючкова Т.Б. Язык и идеология (К вопросу об отражении идеологии в языке). Автореф. канд. дисс. М., 1976.
- Щувалов В.И. Семантические процессы меллорации. — Вопросы немецкой филологии. Сборник трудов. М., 1975.
- Fredriksson, G. Det politiska språket. Uddevalla, 1969.
- Klaus, G. Sprache der Politik. Berlin, 1971.
- Kleivane, S. Reklamespråkets argumentasjon. Oslo-Bergen-Tromsø, 1974.
- Standard Dictionary of the English Language. Part 1. N.Y., 1963.



ON THE EVALUATIVE COLOURING OF  
POLITICAL VOCABULARY

(On the basis of editorials in Dagens Nyheter)

I. Mullaamaa

S u m m a r y

The article deals with the semantic aspect of evaluation in the editorials of the Swedish daily newspaper Dagens Nyheter. The material analysed dates from 1976. Evaluation of political and other events which the newspaper is leading on is considered to be the primary function of the political vocabulary in the functional style under discussion.

The use of political terminology as well as other political vocabulary revealed some typical trends of semantic change. Thus, part of political terminology loses its primary nominative meaning and acquires a secondary evaluative meaning, the terms becoming labels for expressing the author's positive or negative attitude. Some political terms acquire an evaluative meaning that is not fixed and depends on the author's social and political bias. Some typical cases of narrowing and widening of the lexical meaning, contextually imposed meanings as well as melioration and begging the question are discussed.

## SOME NOTES ON RECENT AND CURRENT RESEARCH INTO ENGLISH

### AND THE TEACHING OF ENGLISH (3)

O. Mutt

In recent years the study of language has been characterized by increasing teamwork between scholars and the steadily spreading use of computer techniques for more rapid and efficient processing of linguistic data.

In the field of English studies the past fifteen years or so have seen the launching of various projects aimed at giving a fuller and better description of the vocabulary and grammar of present-day English. In order to achieve this goal large bodies of written and/or recorded spoken text, so-called corpora, have been (or are being) stored on magnetic tape for processing by computer.

The most comprehensive and well-known project of this type is, of course, the Survey of English Usage (S.E.U.) directed by Randolph Quirk since it was started at University College, London, in 1960 (Mutt, 1971: 55-57). With its analysis of samples of the traditional type of printed matter as well as of unprepared speech and free conversation the S.E.U. has already profoundly affected the course of research into present-day English and English language teaching. Large numbers of Anglicists from all over the world have profited from study visits and shorter periods of study attachment to the S.E.U. The publications based on the S.E.U. include the monumental "A Grammar of Contemporary English" (1972) and scores of other books and articles. The grammar has been rightly described as a milestone in the history of English grammar (BIB, 1974: 83). It discloses many so far unnoticed structural relations and meanings in English. The book is not so much a work of reference as a comprehensive new picture of how the grammar of contemporary English actually functions. As such it should be read in full and studied by every serious student of the English language.

Besides the corpus of the S.E.U. the best-known corpus of English available on computer magnetic tape at the present

time is the million-word "Standard Sample of Present-Day Edited American English, for Use with Digital Computers" compiled by a team under W. Nelson Francis at Brown University (U.S.A.) in 1961-4 (Mutt, 1971: 55). Well over a hundred copies of the "Brown Corpus", as it is familiarly known, have been distributed to universities and research institutions throughout the world.

A British counterpart of this American corpus has recently been completed at Lancaster University (U.K.) (Leech & Leonard, 1974). Like the Brown Corpus, the Lancaster Corpus consists of 500 texts of approximately 2000 words each. The corpus is subdivided (again like the Brown Corpus) into genre categories, so as to be representative of varied forms of printed English. A total of fifteen categories have been distinguished (under the headings of Expository Prose and Imaginative Prose), and they include press reportage, press editorials, belles lettres, learned (e.g. scientific, educational and medical) literature, science fiction, mystery and detective fiction, humour, etc. The Brown Corpus has also been followed in taking the texts from printed matter published in the year 1961, and in using random sampling in the selection of texts within the categories.

Although the Lancaster Corpus has been modelled on the Brown Corpus, it is not an "identical twin" of the latter. Cultural differences inevitably lead to differences in subject-matter and the distribution of genre sub-categories, etc. Thus, for instance, reports on cricket matches occur in British newspapers, whereas baseball games are reported on in American papers; fiction dealing with the "Wild West" is published and read on a much smaller scale in the United Kingdom. Despite certain unavoidable minor differences the two corpora are nevertheless broadly comparable and represent a carefully proportioned cross-section of the two principal written varieties of the English language.

The corpora can be used together to provide information on the English language in general (together they make up a combined corpus of 2 million words) or they can be used separately. The existence of these two corpora is of immense value for all Anglicists as one now has, for the first time, a reliable means of testing supposed syntactic and

lexico-syntactic differences between American and British English (e.g. in the use of modal verbs, of the subjunctive, of singular verbs with collective nouns). Contrastive work with the two corpora will no doubt bring to light some less obvious differences which have escaped casual observation.

It should be pointed out, however, that while the Lancaster Corpus exists in an unanalyzed form, the information that can be obtained must be based on coded orthographic symbols and their distribution. The statistical processing of character-sequencing can yield word-frequency lists, and tables and graphs of sentence length of the kind already published for the Brown Corpus (e.g. H. Kučera and W.N. Francis, Computational Analysis of Present-Day American English, 1967). A fairly elementary sorting program can produce a print-out of all occurrences of a particular word or word-sequences, together with their location references and contexts of occurrence. Thus, for example, if a researcher wants to compare the uses of who, whom and whose, he calls for a contextualized print-out of these words in the corpus; if he wants to study adverbs in -ly, he calls for a print-out of all occurrences of ly followed by a space.

Much valuable work can be done with the aid of the Lancaster Corpus in morphological, orthographic and stylistic research and in the lexico-grammatical field. It is possible to compile a frequency-based word-list of the commonest English words, subclassifying the occurrences of words syntactically, listing syntactic and semantic usages in order of frequency and stating the most frequent collocations. For instance, in the case of the following three constructions with the verb think: I thought she was beautiful, I thought her beautiful, and I thought her to be beautiful, one would like to know (for language teaching purposes) which is the most frequent and what factors (stylistic, semantic, syntactic, or collocational) determine the choice of one construction rather than another (Leach & Leonard, 1974: 54).

For the investigation of more abstract grammatical topics, such as word classes, word order, categories such as number, tense, etc. the corpus without prior processing cannot normally yield the required information except by the

time-wasting method of making separate inquiries for each item in a relevant grammatical category. Therefore for most research on grammar as well as on semantics and discourse analysis the corpus must be processed in such a manner that inquiries can be made about whole classes of words and word-sequences. Work is under way at Brown University on an automatic tagging project with the object of producing a grammatically analyzed version of the Brown Corpus. The English Department at the University of Göteborg (Sweden) under A. Ellegård is also working on a tagging project connected with the Brown Corpus. A similar grammatical tagging should be carried out for the Lancaster Corpus, but lack of funds has not made this possible so far (Leech & Leonard, 1974: 53).

The number of linguists interested in linguistic variation and the formulation of rules governing variation has increased rapidly during the past ten years or so (Jacobson, 1976: 8). This interest in variability in language can be traced back to the studies on linguistic competence, acceptability and appropriateness undertaken in the late 1960s, and which were stimulated by the transformational-generative approach to grammar (Mutt, 1970). More recently, too, sociolinguistic studies have encouraged research into language variability.

In 1976 the English Department of Stockholm University started a research project with the purpose of investigating syntactic variation in English (Rydén, 1977: 138). In view of the fact that this field is very comprehensive, the Stockholm project is concentrating on one particular aspect, viz. the use and relative frequency of syntactic variants in different environments, taking account of structure, text, subject-matter, style, situations, etc. The aim is to give as accurate and detailed a picture as possible of the factors governing the selection of variants when there is a possibility of placing adverbs or other sentence elements in different positions, when the use of a particular sentence element is optional (as in The house (that/which) John built), etc. (for other similar cases see, e.g., Leech & Svartvik, 1975: 128 et passim). Frequency judgement tests (in combination with acceptability judgement tests) yield interesting results in comparing cases such as

- a. They insist that she leave the country.
- b. They insist that she leaves the country.
- c. They insist that she should leave the country.

Sidney Greenbaum has found that in these cases the subjunctive is judged to be the most frequent variant but its dominance over the other two variants is much greater after insist than after recommend. S. Greenbaum's informants were Americans, with a British group he would have obtained different results (Jacobson, 1976: 3).

The Stockholm project "Syntactic Variation in English" will be based on a corpus of some three million words drawn from British and American fictional and non-fictional prose, including newspaper texts. It is stipulated that the authors of these texts should have been born not before 1925 and their works published not earlier than 1955. In case of need, the corpus will be supplemented with material obtained from informants (Rydén, 1977: 140). The project will be carried out with the help of contributors at the undergraduate, predoctoral, doctoral and postdoctoral levels.

Two categories of environmental factors are distinguished in the Stockholm project (Jacobson, 1976: 4). The internal environment includes factors belonging to syntax (type of clause, verb, syntactic function, etc.), morphology (contraction, e.g. 'll, n't), phonology (intonation, stress, euphony, rhythm, etc.), graphonomy (punctuation, capitalization), lexicon and semantics. The external environment comprises factors belonging to region (dialect), education and social standing (standard/substandard language), age (child language), medium (written/spoken language), style (formal, colloquial), text category (fiction/non-fiction), subject-matter (scientific, legal English, etc.), preparation (premeditated/spontaneous speech), and idiosyncrasies (individual variation). The influence of various pragmatic phenomena is also being taken into account, i.e. the speaker's or writer's relationship or attitude to the hearer or reader (e.g. adult-child relationship, attitude of deference or superiority), situation (e.g. hurry, tension, dream), etc. (Jacobson, 1976: 6).

Studies of variation are obviously of great pedagogical value. Such work will show which variant in a particular

case is of the greatest importance to learners of English and therefore should be specially emphasized in textbooks and grammars. At an advanced stage of language learning the preferable order of acquiring variants is usually their relative frequency. Corpus investigation and informant tests may even reveal that variants which have been neglected or condemned are actually used by many educated native speakers or writers (Rydén, 1977: 140).

Another Swedish project of great interest is Engelskt Falspråk, the Survey of Spoken English (S.S.E.), launched at Lund University a few years ago. This project is directed by Jan Svartvik, who was for several years associated with the S.E.U. in London, and is, among other things, co-author of "A Grammar of Contemporary English". The S.S.E. is working in close collaboration with the S.E.U. in London and is computerizing a skillfully edited version of the description of the S.E.U. corpus. When the resulting magnetic tapes become generally available, the S.S.E. will have made a substantial contribution to the advanced study of spoken English. In addition there is reason to expect that the computational analysis of English syntax in Lund itself will make outstanding progress in the years ahead (Survey of English Usage Annual Report for 1975-6: 2).

In this brief survey we have referred to some of the major projects under way which will significantly expand our knowledge of English and of how it functions. There are still innumerable facts to be discovered, even about a language as well investigated as English, concerning the nature of the different kinds of English that are used in different situations: who can say what, how, using what means, to whom, when and why? (Crystal, 1974: 254). Extensive work is likewise being conducted both in the Soviet Union and abroad in the fields of English text linguistics, discourse analysis, lexicography and contrastive studies. These are all subjects in their own right and will be dealt with in forthcoming issues of this publication.

## R e f e r e n c e s

- Corder, S. Pit *Introducing Applied Linguistics*. Harmondsworth, England. Penguin Books Ltd., 1973.
- Crystal, D. *Linguistics*. Harmondsworth, England. Penguin Books Ltd., 1974.
- Jacobson, S. *Introduction to the Project Syntactic Variation in English. Part II: For Contributors at the D 1, Doctoral and Post-Doctoral Levels*. Department of English. Stockholm University, 1976.
- Leech, G. & Leonard, E. *A Computer Corpus of British English*. - "Hamburger Phonetische Beiträge" 13 (1974), 41-57.
- Leech, G. & Svartvik, J. *A Communicative Grammar of English*. London. Longman, 1975.
- Matt, O. *Some Problems of Acceptability in Language and the Teaching of English*. - "Linguistics" II. Tartu, 1970, 5-18.
- Matt, O. *Some Notes on Recent and Current Research into English and the Teaching of English*. 1. - "Linguistics" III. Tartu, 1971, 51-61.
- Quirk, R., Greenbaum, S., Leech, G., Svartvik, J. *A Grammar of Contemporary English*. - Review in "English Language Teaching" Vol. XXIX (1974), No. 1, 83-88.
- Rydén, M. *Introduction to the Project Syntactic Variation in English. Part I: For Contributors at the 04 Level*. Department of English. Stockholm University, 1976.
- Rydén, M. *Syntaktisk variation in engelskan: Presentation av ett projekt*. - "Papers from the Conference on Contrastive Linguistics and Error Analysis. Stockholm & Åbo 7-8 February 1977. Meddelanden från stiftelsen för Åbo Akademi Forskningsinstitut, nr. 19. Åbo 1977, 137-140.
- Survey of English Usage. Annual Report 1975-6. Mimeo. University College London. London, 1976.*



НЕКОТОРЫЕ ЗАМЕТКИ О СОВРЕМЕННЫХ ИССЛЕДОВАНИЯХ  
В ОБЛАСТИ АНГЛИЙСКОГО ЯЗЫКА И ЕГО ПРЕПОДАВАНИЯ. III

О. Мутт

Р е з ю м е

В статье дается обзор целей, методики и результатов некоторых более крупных зарубежных коллективных исследований в области англистики. В частности рассматривается современное состояние работы над *Survey of English Usage* (Лондонский университет), и работа, проводимая при английском университете в Ланкастере, а также при шведских университетах в Стокгольме (вариантность в английском синтаксисе) и Лунде (английская разговорная речь).

## МЕТОД ПОЛНОЙ ФИКСАЦИИ ТЕКСТА ПРИ ЛИНГВОСТАТИСТИЧЕСКОМ АНАЛИЗЕ

И.Ш. Надарейшвили, Ю.К. Орлов.

### I.

Количественные методы в исследовании языка и речи стали в настоящее время обычным явлением, множество фактических данных уже опубликовано, защищено немалое число диссертаций (в т.ч. и таких, приложение к которым с изложением фактических данных выполнено на ЭВМ) - и тем не менее попытки как-то обобщить имеющийся материал или применить для его обработки математические методы встречаются с немалыми затруднениями. Причин этому две - отсутствие унификации при подсчетах и существенная неполнота публикуемых обычно данных. Очень часто отсутствуют даже данные о частотной структуре (список всех наблюдаемых частот слов: употребляются также термины "лексический спектр" (Калинин, 1964), "лингвистический спектр" (Алексеев, 1975, и др.)). Вот что пишет об этой наблюдавшейся проблеме П.М. Алексеев (1975): "К сожалению, такие таблицы в полном виде приводятся не в каждом частотном словаре; более того, их нельзя найти почти нигде, за редким исключением / ... /. Составителю, проделавшему длительную и утомительную работу над словарем, как будто не хватает терпения потратить еще несколько часов на изготовление таблицы. В конце концов читатель может составить такую таблицу и сам, если словарь содержит хотя бы основные данные: частоты и количества слов с одинаковыми частотами. Однако, / ... /, в публикуемый словарь обычно включают ограниченное число входящих в него единиц. Если об оставшихся за пределами публикации единицах не даются количественные сведения, такой словарь во многом теряет свою лингвистическую ценность. Чтобы сопроводить такой словарь таблицей, где указывалось бы количество разных единиц на каждую частоту, потребуется время ничтожное по сравнению с тем, которое было затрачено на составление словаря".

Одна из причин этого явления связана с тем, что в публикацию обычно выносятся только те данные, которые так или

иначе работают на авторскую концепцию (ради которой и проводились подсчеты). Не имеющие к ней отношения цифры (которые неизбежно должны были быть получены в процессе подсчетов) попросту опускаются — и это делает результаты подсчетов недоступными для других исследователей. И это несмотря на то, что лингвистические подсчеты чрезвычайно трудоемки (даже когда они выполняются с помощью ЭВМ) и поэтому результаты этих подсчетов представляют некоторую безусловную ценность, совершенно не зависящую от той концепции, которую пытается обосновать с их помощью автор подсчетов. Лингвостатистика не может позволить себе роскошь подражания, например, физике, в которой выполняются лишь измерения, необходимые для проверки той или иной гипотезы. Лингвостатистика (да и вся вычислительная лингвистика) только-только приступила к накоплению фактов — и поэтому очень важно, чтобы каждый исследованный факт был зафиксирован не только точно и тщательно, но и допускал бы сравнение с другими фактами, исследованными другими людьми в другом месте.

Оставляя в стороне чисто лингвистические проблемы унификации подсчетов (например, проблему определения слова или различения омонимов), в данной работе остановимся на проблеме полноты количественного описания лингвистического объекта.

Что можно понимать под полным описанием некоторого объекта? Очевидно, такое описание, по которому, хотя бы в принципе, можно восстановить сам объект. Обычный же лингвостатистический анализ приводит к необратимой деструктуризации объекта — очевидно, что по частотному словарю некоторого текста совершенно невозможно восстановить сам текст. В принципе текст может быть восстановлен по индексу (или конкордансу), в котором указываются адреса (страница и строка в определенном издании текста) употреблений каждого слова. Следовательно, именно индекс является полным описанием текста. Однако, с точки зрения количественного анализа, индекс является лишь чрезвычайно сырым полуфабрикатом — нужны очень большие усилия, чтобы извлечь из него необходимые цифры.

Традиционная форма частотного словаря обладает еще одним существенным недостатком. Она фиксирует лишь некоторую статистику текста, оставляя в стороне вопросы динамики. Например, когда хотят исследовать нарастание словарного за-

паса с ростом объема текста, приходится организовывать дополнительные громоздкие и чреватые ошибками процедуры разбиения текста на подвыборки, пересчет подвыборок и последующее объединение словарей подвыборок. Объемы подвыборок при этом определяют степень подробности, с которой мы можем изучить процесс нарастания словарного запаса. Если, например, текст был разбит на подвыборки по 1000 словоупотреблений, то после объединения этих подвыборок мы будем знать словарь текста только через каждую тысячу словоупотреблений. При необходимости изучить рост словаря более детально (например, через каждую сотню словоупотреблений), старые подсчеты оказываются непригодными и счет нужно начинать заново. Практически, конечно, так никто никогда не поступает - приходится довольствоваться тем разбиением, которое было принято в начале работы на основании априорных соображений, хотя оно, возможно, было далеко от оптимального для решения поставленной задачи. И уж наверняка оно не будет удачным для других исследователей, желающих воспользоваться опубликованными данными.

В настоящей работе описывается метод полной фиксации текста при лингвостатистическом анализе, позволяющий сохранить информацию как о статике, так и о динамике лингвистических единиц в тексте с любой степенью подробности. По общей трудоемкости этот метод лишь незначительно сложнее обычной процедуры составления частотного словаря. Метод описывается в применении к обычной технике работы с карточками; машинная его модификация не представляет проблемы.

## 2.

После того, как выбрана единица счета (например, решено просчитывать текст на уровне лексем; но можно выбрать и любую другую единицу), текст расписывается на заранее определенные номера на карточки. Поскольку процедура нумерации карточек трудоемка и ответственна (не обнаруженные своевременно - т.е. до начала расписания - ошибки нумерации не поддаются исправлению), нужно воспользоваться имеющимся в продаже полуавтоматическим нумератором типа НП. Нумерация начинается с единицы и каждое словоупотребление оказывается, т.о., скрепленным с номером своей позиции в тексте.

После обычного библиографического упорядочения (по алфавиту) все карточки, на которых зафиксировано одно и то же слово, оказываются вместе. Если упорядочение было произведено аккуратно, то номера позиций, на которых данное слово встретилось в тексте, также оказываются упорядоченными либо в прямом, либо в обратном порядке (в зависимости от числа перекладок). Эти карточки пересчитываются и для словарной картотеки отбирается карточка с наименьшим номером позиции, фиксирующим первое вхождение данного слова в текст. На нее записывается полное число всех вхождений данного слова в текст (частота), а также (на обороте карточки: если слово очень частое, то можно подклеить дополнительную полоску бумаги и сложить ее гармошкой) номера всех остальных позиций вхождения данного слова в текст (в возрастающем порядке). Полученный алфавитно-частотный словарь фиксируется; он оказывается аналогом обычного индекса. Разница в том, что по номеру позиции труднее отыскать заданное слово в тексте (для этого нужно перенумеровать слова в тексте), однако есть и преимущество: в то время, как для работы с индексом необходимо располагать совершенно конкретным изданием текста, работа со словарем, в котором зафиксированы номера позиций, возможна с любым изданием данного текста. Следует заметить, что разница в трудоемкости использования обоих видов словарей может почти исчезнуть, если во время расписывания на карточки фиксировался номер позиции первого на каждой следующей странице слова. Располагая таким списком, легко отыскать страницу; остальное не составляет большого труда. Разумеется, как и в случае обычного индекса, такой список действителен только для того издания текста, по которому производилось расписывание карточек.

После того, как алфавитный список зафиксирован, словарная картотека может быть упорядочена по возрастанию номеров позиций первого вхождения слов в текст. Слова при этом выстраиваются в порядке их первого появления в тексте. Нетрудно догадаться, что такой список содержит предельно подробную информацию о нарастании количества разных слов с ростом объема текста. Например, слово ГЕРМАНН, впервые встретившееся в "Пиковой даме" А.С. Пушкина на 130-м месте, в списке слов, упорядоченных в порядке первого появления, находится на 98-м месте. Это значит, что на первых 130 словоупотреблениях тек-

ста накопился словарь в 98 слов. Т.о., процесс нарастания словарного запаса по такому списку прослеживается с точностью до одного слова. Следует отметить, что процедура перетасовки карточек гораздо менее трудоемка и менее чревата ошибками (а допускаемые ошибки видны и поддаются исправлению), чем процедура объединения и пересчета словарей подвыборок. Фиксируя перетасованную картотеку, нужно записать частоту каждого слова и номер первого вхождения слова (см. Приложение I); номера позиций всех остальных вхождений слова в текст (если это не требуется некоторыми специальными целями исследования) фиксировать уже не нужно, поскольку они зафиксированы в алфавитном списке.

В качестве иллюстрации возможностей работы с таким списком приведем анализ участка текста из "Пиковой дамы", содержащий два смежных эпизода: томительное ожидание Германа после отъезда Графини и ее воспитанницы (от слов "Дверцы захлопнулись" до слов "Лампа слабо освещала их из передней") и его впечатления в спальне Графини (от слов "Герман вошел в спальню" до слов "... вместе с Монгольфьеровым шаром и Месмеровым магнетизмом."). Каждый из эпизодов содержит 101 словоупотребление; их позиции в тексте — от 3420 по 3621. В Приложении I приведен участок словарного списка, относящийся к этому участку текста. Горизонтальная черта отделяет слова, относящиеся к первому эпизоду, от слов второго эпизода.

Легко видеть, что в первом эпизоде употреблено 26 разных слов, каждое из которых не употреблялось в тексте ранее. 18 из них употреблены во всем тексте ровно 1 раз (цифры в скобках) — т.е. они употреблены лишь в этом эпизоде. Во втором же эпизоде употреблено 56 разных слов и 44 из них — по одному разу во всем тексте. Иными словами, лексическая насыщенность обоих эпизодов как новыми, ранее не употреблявшимися словами, так и словами, употребленными только в этом эпизоде, существенно различна. Невольно закрадывается мысль — а не является ли подобная дозировка словарного запаса своего рода средством художественной выразительности, воздействующим на читателя помимо его сознания? Мог же ведь Пушкин, в конце концов, найти свежие слова для описания окружающих Германа предметов, пока тот ждет условленного часа (как это нередко делают писатели меньшего калибра)?

Подобная же дозировка в несравненно больших масштабах наблюдается в начальной части "Казakov" Л.Н. Толстого. В ра-

боте И. Надарейшвили, Ю. Орлова (1971) приводится график нарастания словарного запаса на первых 10000 словоупотреблений этого текста. Из этого графика видно, что на первых 4000 словоупотреблений рост словаря несколько замедленный (во всяком случае, по сравнению с теоретической кривой, приведенной на том же графике). Этот участок текста описывает путешествие Оленина на Кавказ — однообразная белая равнина, глухие городишки похожи друг на друга, сам Оленин дремлет в дороге. Но вот он приезжает на Кавказ — и мир вспыхивает ослепительными красками. Эти первые кавказские впечатления соответствуют участку крутого подъема кривой роста словаря (5-я — 6-я тысяча словоупотреблений), после чего ход кривой выравнивается и она начинает совпадать со "своей" теоретической кривой (подробнее количественный анализ этой ситуации рассмотрен в Приложении 2).

Таким образом, упорядочение словарной картотеки в той последовательности, в которой слова впервые появляются в тексте, позволяет не только с определенной степенью подробности описать чисто количественную сторону процесса нарастания словарного запаса с ростом текста, но и дает возможность ставить и решать филологические и лингвистические проблемы. Словарь такого типа чрезвычайно легко "привязывается" к тексту и с его помощью можно легко решать такие проблемы, как место новых (для данного текста) слов в предложении, распределение их по позициям стихотворного текста и т.п.

Техника нумерации карточек и выделение позиции первого вхождения слова в текст предоставляет новые возможности и при организации частотного списка лексики. Если словарную картотеку, упорядоченную в порядке первого появления слов, разложить по убыванию частоты, то все слова с одинаковой частотой "сами собой" (если раскладка производилась аккуратно) расположатся также в порядке их первого появления в тексте. Т.е. взамен обычно практикуемого алфавитного упорядочения слов одной частоты, не имеющего филологического или лингвистического смысла, возникает упорядочение, диктуемое самим исследуемым объектом и отражающее особенности его структуры. Такому упорядочению подвергнутся все редкие слова, составляющие львиную долю словаря текста. В зоне же частых слов представляют интерес связи между частотой слова и номером позиции его первого появления. Естественно, что наиболее частые слова должны впервые встретиться сравнительно рано

(т.е. номера позиций их первого вхождения в текст будут малы). Если же частое слово будет иметь высокий номер позиции первого вхождения в текст, это будет означать, что это слово специфично лишь для некоторой части текста. Ранее зафиксированные в алфавитном списке номера всех позиций вхождения слова позволят точно определить особенности и границы его употребления; в частности, легко могут быть решены задачи типа решаемых в работах: Токарев, Якубайтис, 1969; Бектаев, Лукьяненко, 1971; Каширина, 1974; при этом появляется возможность выбора любого разбиения текста на подвыборки и сравнения результатов, полученных при разных разбиениях.

С точки зрения количественного анализа организованный таким образом частотный словарь позволяет не только исследовать динамику появления слов той или иной частоты, но и, при необходимости, извлечь информацию о частотной структуре любого отрывка из текста. Излишне говорить, что обычный частотный словарь этого не позволяет в принципе. Возможно также исследование позиций слов малой частоты (см. Борода, Надарейшвили, Орлов, Читашвили; Г. Надарейшвили, И. Надарейшвили, Орлов, 1975) и многое другое.

### 3.

На первый взгляд кажется, что применение техники нумерации карточек имеет смысл лишь при анализе отдельных текстов, когда позиция каждого слова не зависит от произвола исследователя. Однако исследование кривой роста словарного запаса внутри корпуса различных текстов, выявление изломов кривой на границах между текстами, позволяют сделать обоснованные заключения о степени статистической однородности исследуемого корпуса (см. Приложение 2, а также Орлов, 1977; Нешитой, 1976; Тулдава, 1971). Если при расписывании карточек были зафиксированы номера позиций, соответствующие границам перехода от текста к тексту, то из описываемого словаря становится возможным извлекать и ту информацию, которую содержит т.н. распределительный словарь (см. Андреев, 1965). Особый интерес применение подобной техники приобретает тогда, когда корпус документов можно упорядочить хронологически. В этом случае, в сущности, лингвистическими методами начинает исследоваться некоторый процесс нелингвистической природы, про-



текающий во времени<sup>II</sup>.

Предлагаемая форма фиксации результатов лингвистических подсчетов в виде трех словарных списков:

1) алфавитного с указанием частоты и номеров всех позиций встречи данного слова, упорядоченных по возрастанию,

2) списка слов, упорядоченных в порядке первого появления в тексте (с указанием частоты) и

3) частотного списка (с указанием номера первого вхождения слова в текст и с упорядочением слов, имеющих одну частоту, в порядке возрастания этого номера).

дает в удобном и обозримом виде полную информацию о тексте. Эта информация пригодна для самого разнообразного анализа, как количественного, так и традиционно-филологического. Нет смысла спорить, какой из видов словарей — индекс, частотный или распределительный — лучше. При составлении любого из них производятся одни и те же, чрезвычайно трудоемкие операции. И поэтому очень важно зафиксировать результаты этих операций в такой форме, чтобы ничего не пропало, чтобы результатами подсчетов мог воспользоваться не только их автор (преследующий, как правило, достаточно узкую цель), но и другие исследователи. Количественная, вычислительная лингвистика делает пока только первые шаги; время широких обобщений и глубоких гипотез для нее еще вне ради. Но, чтобы это время пришло, нужно накапливать фактический материал, нужно, чтобы делаемое сегодня не пропадало для будущего, чтобы результаты подсчетов фиксировались с той же полнотой и тщательностью, которая издавна отличала классическую филологию.

Учитывая трудности с опубликованием трех словарных списков, было бы разумно наладить практику депонирования таких работ и издания реферативных сборников. В реферат должны входить основные параметры текста — объем текста и объем его словаря, а также полный список всех частот (частотный спектр) и (желательно) список наиболее частых слов, а также сведения о нарастании словарного запаса с ростом объема текста (например, указать объем словаря через каждую 1000 или

<sup>II</sup> Эта мысль о возможностях описываемого метода для исследования динамики информационных потоков родилась во время беседы одного из авторов данной работы с Г.Ф. Крайчинской (Патентный отдел УКРНИИПЛАСТАМАШ, г. Киев) о проблеме анализа текстов описаний изобретений.

10000 словоупотреблений). Депонирован должен быть, как минимум, алфавитный список с указанием номеров всех позиций; желательно же, конечно, депонировать все три списка и приложить также сведения о номерах позиций первых слов на каждой странице текста, о разбеге текста на главы, книги, тома и т.п., а если обобщивался корпус документов — о границах между отдельными документами. Подобная практика могла бы существенно ускорить развитие лингвостатистики.

Авторы благодарны доктору филологических наук В.С. Перейносу и сотрудникам возглавляемого ею отдела структурно-математической лингвистики Института языкознания АН УССР за подробное обсуждение предлагаемого метода фиксации результатов лингвистических подсчетов.

#### Приложение I.

Приводим участок словаря "Пиковой дамы", относящийся к разбираемому в тексте эпизоду. Слова упорядочены в порядке первого (от начала текста) появления. Номер позиции первого появления — правая колонка. Порядковый номер слова, зафиксированный в левой колонке, одновременно является объемом словаря, накопившимся к моменту появления данного слова. В частности, словарь, накопившийся в тексте "Пиковой дамы" до начала первого эпизода, равен 1167 словам, к концу второго эпизода, на объеме 3621 словоупотреблений — 1250 словам. В круглых скобках дана абсолютная частота слова в тексте.

№ п/п	№ поз.	№ п/п	№ поз.
II68 захлопнуться (1)	3423	II82 взойти (2)	3473
II69 тяжело (1)	3425	II83 ярко (1)	3475
II70 покатиться (2)	3426	II84 взбежать (1)	3482
II71 запретить (1)	3430	II85 спящий (1)	3492
II72 померкнуть (1)	3433	II86 лампа (9)	3494
II73 опустевший (1)	3438	II87 запачканный (1)	3497
II74 часы (5)	3446	II88 кресло (1)	3498
II75 двадцать (2)	3448	II89 легкий (1)	3499
II76 часовой (1)	3457	II90 твердый (2)	3501
II77 стрелка (1)	3458	II91 шаг (2)	3502
II78 выкидать (1)	3460	II92 темный (3)	3511
II79 остальной (1)	3461	II93 слабо (1)	3513
II80 ровно (2)	3463	II94 освещать (1)	3514
II81 ступить (1)	3468	II95 кивот (1)	3526

№ п/п	№ поз.	№ п/п	№ поз.
И196 наполненный (2)	3527	И224 орлиный (I)	3581
И197 образ (I)	3529	И225 нос (I)	3582
И198 теплится (I)	3530	И226 зачесанный (I)	3584
И199 золотой (I)	3531	И227 висок (I)	3585
И200 лампада (I)	3532	И228 роза (2)	3588
И201 полинялый (I)	3533	И229 пудренный (I)	3590
И202 штофный (2)	3534	И230 волосы (I)	3591
И203 диван (3)	3537	И231 торчать (I)	3595
И204 пуховый (I)	3539	И232 фарфоровый (I)	3596
И205 подушка (I)	3540	И233 пастушок (I)	3597
И206 сошедший (I)	3542	И234 столовый (I)	3598
И207 позолота (I)	3543	И235 славный (3)	3601
И208 печальный (2)	3546	И236 Legoу (I)	3602
И209 симметрия (I)	3547	И237 коробочка (I)	3603
И210 стена (2)	3549	И238 рулетка (I)	3604
И211 обитый (I)	3550	И239 веер (I)	3605
И212 китайский (I)	3551	И240 разный (I)	3607
И213 висеть (I)	3555	И241 дамский (I)	3608
И214 портрет (3)	3557	И242 игрушка (I)	3609
И215 писанный (I)	3558	И243 изобретенный (I)	3610
И216 М-ше (I)	3561	И244 конец (I)	3612
И217 Lebгun (I)	3562	И245 минувший (I)	3613
И218 изображать (2)	3566	И246 столетие (I)	3614
И219 сорок (3)	3569	И247 Монгольферыов (I)	3617
И220 румяный (I)	3570	И248 шар (I)	3618
И221 светлозеленый (I)	3572	И249 Месмеров (I)	3620
И222 мундир (I)	3573	И250 магнетизм (I)	3621
И223 звезда (I)	3576		

Просматривая приведенный список, легко заметить еще одно преимущество подобной организации словаря. Это единственный из п о с л о в н ы х словарей, позволяющих выявлять некоторые типовые словосочетания, такие, как "твердый шаг", "золотая лампада", "дамская игрушка", "пуховая подушка", "орлиный нос" и др. Причем это выявление происходит независимо от частоты данного сочетания.

# АНАЛИЗ КРИВЫХ РОСТА СЛОВАРЯ

Упорядочение словаря в порядке первого появления слов в тексте дает возможность с предельной детальностью, буквально пословно изучать процесс нарастания словаря с ростом объема текста. Если такие данные будут систематически публиковаться (или, по крайней мере, депонироваться), со временем это откроет широкое поле для сравнительного анализа, при этом полностью отпадает сегодняшняя проблема несопоставимости результатов (одна выборка содержит 5000 словоупотреблений, другая 25500, а третья 101000). Однако уже сегодня возможен содержательный теоретический анализ зависимостей словарного запаса от длины текста путем сравнения реально наблюдаемых данных с теми или иными теоретическими моделями.

Среди всех предложенных способов построения теоретических кривых роста словарного запаса с ростом объема выборки следует особо выделить формулу В.М. Калинина (1964), имеющую вид:

$$E_v(N) = E_v(N_0) - \sum_{j=1}^{\infty} \left(1 - \frac{N}{N_0}\right)^j E_{vj}(N_0) \quad (1)$$

Здесь  $E_v(N_0)$  - мат. ожидание значения словарного запаса на случайной выборке объемом  $N_0$  словоупотреблений;  $E_{vj}(N_0)$  - мат. ожидание числа  $j$  - разовых слов на той же выборке;  $E_v(N)$  - мат. ожидание словарного запаса на выборке произвольного объема  $N$ . Поскольку мат. ожидания  $E_v(N_0)$  и  $E_{vj}(N_0)$  обычно нам неизвестны ни для какого объема  $N_0$  (мы можем лишь строить гипотезы об их виде, подобно тому, как это сделано в работе: (Орлов, 1977)), кажется, что формула (1) не обладает практической ценностью. Однако, при подстановке вместо математических ожиданий  $E_v(N_0)$  и  $E_{vj}(N_0)$  фактически наблюдаемых значений объема словаря  $v^*(N_0)$  и числа  $j$  - разовых слов  $v_j^*(N_0)$  на выборке объема  $N_0$ , эта формула дает точное значение мат. ожиданий  $E_v(N)$  для любой случайной подвыборки\* (т.е.  $N$  должно быть

\* Строго говоря, случайная выборка из конечной совокупности (без возврата и перемешивания) описывается гипергео-

меньше  $N_0$ , иными словами, прогноз от наблюдаемых значений словаря и спектра по формуле (I) возможен только "назад"). В случае  $N > N_0$  прогноз "вперед" по фактически наблюдаемым значениям по формуле (I) недостоверен.

Формула Калинина пригодна для описания процесса нарастания числа любых единиц, поскольку она не зависит от их распределения. Любое отклонение фактически наблюдаемой кривой роста словаря от формулы (I) следует трактовать как следствие неслучайности подвыборки — иными словами, при анализе текста единственной причиной расхождения между

метрическим распределением (см., напр., Большев, Смирнов, 1965, II4), которое приводит к следующим зависимостям для частотного спектра и словаря:

$$E_{v_m}(N) = \sum_{k=m}^{N_0-N-m} v_k^*(N_0) \frac{\binom{k}{k-m} \binom{N_0-k}{N_0-m}}{\binom{N_0}{N}}$$

$$\begin{aligned} E_v(N) &= \sum_{k \geq 1} E_{v_k}(N) = \sum_{k \geq 1} v_k^*(N_0) \sum_{i=\max(k-N, 0)}^{\min(N_0-N, k-1)} \frac{\binom{k}{i} \binom{N_0-k}{N_0-N-i}}{\binom{N_0}{N}} = \\ &= \sum_{k=1}^{N_0} v_k^*(N_0) \left(1 - \frac{\binom{N_0-k}{N_0-N-k}}{\binom{N_0}{N}}\right) - \sum_{k \geq N+1} v_k^*(N_0) \sum_{i=0}^{k-N-1} \frac{\binom{k}{i} \binom{N_0-k}{N_0-N-k}}{\binom{N_0}{N}} \end{aligned} \quad (1a)$$

Если  $N$  превышает абсолютную частоту наиболее частого слова на объеме  $N_0$ , то последняя сумма обращается в нуль и

$$E_v(N) = v^*(N_0) - \sum_{k \geq 1} v_k^*(N_0) \frac{(N_0-N-k+1)(N_0-N-k+2) \dots (N_0-N)}{(N_0-k+1)(N_0-k+2) \dots N_0} \quad (1b)$$

Однако в практически интересном для лингвостатистики диапазоне значений  $N$  и  $N_0$  разница между приведенными формулами и формулой (I), полученной при допущении справедливости распределения Пуассона, совершенно несущественна. Она становится заметной лишь при очень малых  $N$  (порядка единиц и немногих десятков) из-за возрастания двойной суммы в правой части формулы (1a). Формулы (1a) и (1b) могут использоваться для контроля формулы (I) в сомнительных случаях.

теоретическим и фактическим ростом словаря может быть только статистическая неоднородность текста. Это делает формулу Калинина своего рода эталоном для проверки гипотезы об однородности данного текста (или корпуса текстов).

Формула Калинина описывает рост словарного запаса в идеально бесструктурном тексте. Такой "текст" может быть получен, например, тщательным перемешиванием карточек после расписывания. Если эти карточки перенумеровать в случайном порядке (после перемешивания), то кривая роста словаря в таком "тексте" должна точно совпасть с вычислениями по формуле (I), незначительно отклоняясь от нее в обе стороны в результате случайных флуктуаций.

Любопытно, что в лингвостатистической практике нередко организуют процедуры, результат которых в той или иной степени приближается к вышеописанному перемешиванию. Таковы, в частности, исследования словарного запаса на выборках разного объема в случае заведомо разнородного исходного материала (например, языка газеты), когда в подвыборках тексты из разных рубрик представляются в той же пропорции, что и во всем корпусе материалов. Таковы же т.н. "механические" выборки, когда подвыборка образуется дроблением текста на равные фрагменты, находящиеся на одинаковом расстоянии друг от друга (например, верхняя треть страницы). В сущности, эти процедуры совершенно излишни, поскольку их результат точно предсказывается формулой Калинина. Если все же произойдет расхождение, это будет означать лишь то, что предпринятая процедура "усреднения" подвыборок не достигла своей цели и данная подвыборка не является "репрезентативной". Мы имеем дело со случаем, когда теоретическим вычислениям можно доверять больше, чем экспериментальным данным.

Таким образом, располагая экспериментально наблюдаемым частотным спектром  $\nu_m^*(N)$ ;  $m = 1, 2 \dots$  на выборке некоторого объема  $N_0$  и значениями словарных запасов на входящих в нее подвыборках  $\nu^*(N)$  ( $N < N_0$ ), можно, пользуясь формулой (I), оценить степень однородности подвыборок (для связного текста — степень однородности текста) или, если подвыборки или текст заведомо неоднородны (например, текст разбит на две подвыборки — авторская речь и прямая речь персонажей) — влияние этой неоднородности на словарный запас. Достаточно лишь вычислить по формуле (I) теоретический словарь на объемах подвыборок и сравнить его с наблюдаемым.

Аналогичные по идее процедуры оценки степени однородности предлагались и в других работах (Орлов, 1977; Нешиной, 1976; Тулдава, 1971). Все они базируются на подтверждаемых эмпирически кривых роста словарного запаса (с одним или двумя параметрами), благодаря чему достаточно знать лишь словарные запасы на разных объемах (в отличие от формулы Калинина, которая требует знания частотного спектра выборки). Однако, во всех трех случаях нет никаких гарантий, что наблюдаемые расхождения между эмпирическими и теоретическими словарными запасами вызваны только неоднородностью текста, а не несовершенством самих формул.

Разберем с этой точки зрения работу хорошо зарекомендовавшей себя формулы роста словарного запаса (Надарейшвили, Орлов, 1971; Орлов, 1977):

$$\nu(N, Z) = \nu(Z) \frac{\ell_n \frac{Z}{N}}{\frac{Z}{N} - 1}, \text{ где } \nu(Z) = \frac{Z}{\ell_n(Z\rho_1)} \quad (2)$$

Эта формула была получена в работах Ю. Орлова (1976; 1977) как результат подстановки в формулу Калинина (1) некоторой гипотезы о виде мат. ожиданий частотного спектра на объеме выборки  $Z$ ;  $N$  - текущий объем,  $\nu(N, Z)$  - ожидаемый объем словаря на выборке объема  $N$ ,  $\nu(Z)$  - ожидаемый объем словаря на выборке объема  $Z$ ,  $\rho_1$  - относительная частота самого частого слова в тексте. Параметр  $Z$  может быть оценен из уравнения:

$$\nu(N_0, Z) = \nu^*(N_0) \quad (3)$$

где  $\nu^*(N_0)$  - фактически наблюдаемый объем словаря на выборке объемом  $N_0$  словоупотреблений. Способы решения трансцендентного уравнения (3) даны в работе Ю. Орлова (1977); там же дано построение доверительных интервалов для прогнозов по формуле (2).

В отличие от формулы Калинина (1) формула (2) пригодна для прогнозов не только назад, но и вперед; кроме того, она не требует знания частотной структуры на исходной выборке. Пример десятикратного прогноза как вперед, так и назад по этой формуле приводится в работе Ю. Орлова (1977); тридцатикратный прогноз вперед содержится в работе Ю. Орлова (1976).

В худших случаях ошибки прогноза по формуле (2) не выходят за пределы 10+12%. Однако расхождения между фактически наблюдаемым словарем и прогнозом по формуле (2) могут происходить, в отличие от формулы (1), не только вследствие неоднородности текста, но и вследствие неадекватности (в отношении данного текста) принятой при выводе формулы (2) гипотезы о его частотной структуре. Причем расхождения между теоретическими прогнозами, полученными по формулам (1) и (2), характеризуют именно степень этой неадекватности. Практически, однако, нередко случается так, что при наличии расхождений между формулами (1) и (2), формула (2) лучше описывает фактический рост словаря в тексте (т.е., приближенность идеализации, положенной в основу вывода формулы (2), как бы компенсируется неоднородностью реальных текстов, в частности, явлением сгущивания слов, вызывающим замедленный рост словаря (см. Борода, Надарейшвили, Орлов, Читашвили, в печати).

Продemonстрируем работу формул (1) и (2) на примере роста словаря в "Пиковой даме":

Т а б л и ц а I

Объем выборки	О б ъ е м   с л о в а р я		
	фактич. $\nu^*(N)$	по ф-ле (1)	по ф-ле (2)
500	281	308	293
1000	462	523	497
2000	787	864	845
4000	1348	1371	1360
6000	1752	1778	1770
6861	1928	1928	1930*

Вычисления по формуле (1) производились при подстановке наблюдаемого на полном объеме текста  $N_0 = 6861$  частотного спектра  $\nu_m^*(N_0)$ , фактический словарь на этом же объеме  $\nu^*(N_0) = 1928$  был подставлен и в уравнение (3) для нахождения параметра  $Z$ . При  $p_i = 0,038$  (фактическая частота слова "и" в "Пиковой даме") из этого уравнения находим  $Z \approx 35000$ . Звездочкой отмечено совмещенное с фактическим словарем теоретическое значение словарного запаса в результате решения уравнения (3). Все остальные прогнозы по формуле (2) вычис-



лем при найденном  $Z = 35000$ .

Как видно из табл. 1, почти на всех объемах обе формулы описывают рост словаря в данном тексте вполне удовлетворительно. Однако в начальной части текста (примерно первые 2 главы, до начала интриги Германина) фактический рост словаря несколько замедлен по сравнению с прогнозами по формуле (1). Это может быть объяснено только, как отмечалось выше, неоднородностью связного текста. Прогнозы по формуле (2) несколько ближе к фактическому росту словаря, но и в этом случае ощущается влияние неоднородности текста.

Более "рельефная" картина расхождений между формулами и фактическим ростом словаря наблюдается на первых 10000 словоупотреблений "Казаков" Л.Н. Толстого (табл. 2):

Т а б л и ц а 2

Объем выборки	О б ъ е м   с л о в а р я		
	фактич.	по ф-ле (1)	по ф-ле (2)
500	274	337	298
1000	438	548	513
2000	732	922	862
4000	1233	1472	1395
6000	1814	1904	1840
8000	2253	2266	2225
10000	2582	2582	2600*

В этом случае для расчетов также использовался словарь и спектр на конечном объеме; подстановка  $\nu^*(10000) = 2582$  в уравнение (3) дает (при  $\rho_1 = 0,05$ )  $Z \approx 53000$ . Как видно из табл. 2, расхождения между фактическим ростом словаря в "Казаках" и прогнозами по обоим формулам существенно больше, чем в случае "Пиковой дамы". Это естественно объясняется крайней неоднородностью обследованного отрывка, состоящего из двух крупных контрастных участков (см. основной текст). Как и в предыдущем случае, формула (2) несколько лучше описывает фактический рост словаря, чем формула (1). Данный пример поучителен в том смысле, что из этого факта никак нельзя сделать вывод, что формула Калинина хуже, чем формула (2). Резкое з а н и ж е н и е фактического словарного запаса

по сравнению с ожидаемым для случая многого выбора слова из текста с данной частотной структурой свидетельствует о том, что при выборке подряд от начала текста место многих новых слов занимает "лишними" повторениями слов, уже употребленных (если бы наблюдалось завышение фактического словаря на начальном участке текста, это значило бы, что автор нарочито избегает использования уже употребленных слов в начале текста с тем, чтобы повторить их в конце). То есть "эталонный" характер формулы Калинина позволяет в случае расхождений сделать позитивные утверждения о структуре текста.

Формула (2) в этом случае несколько расходится с формулой Калинина. Это свидетельствует о том, что частотная структура отрывка из "Казakov" слегка отличается от постулированной при выводе формулы (2). Есть основания предполагать, что начальный участок текста подействовал искажающим образом и на частотную структуру обследованного отрывка и что при дальнейшем увеличении объема выборки разница между прогнозами по формулам (1) и (2) уменьшилась бы. Но и на основании сравнения фактического роста словаря с прогнозом по формуле (2) можно заключить о существенной неоднородности текста.

Чрезвычайно любопытно сравнить табл. 2 с аналогичными данными по "Воскресения" Л.Н. Толстого (также первые 10000 словупотреблений от начала текста, табл. 3). Ход прогнозов по формуле (2) не изменился, поскольку практически не изменились положенные в основу расчета параметра исходные данные ( $P_1 = 0,05$ ,  $\psi^*(10000) = 2587$ ). Несколько изменился ход прогнозов по формуле (1), в результате чего прогнозы по формулам (1) и (2) стали очень хорошо совпадать друг с другом: также хорошо совпадают с этими прогнозами фактические значения словарного запаса, что свидетельствует как о высокой однородности текста, так и соответствии частотной структуры текста той гипотезе, которая была принята при выводе формулы (2).

С содержательной точки зрения высокая степень однородности "Воскресения" несколько парадоксальна, по крайней мере на первый взгляд. Начальный участок романа состоит из контрастных эпизодов, весьма разных по лексике, рисующих быт Катюши Масловой в тюрьме и быт князя Нехлюдова. Однако эти эпизоды чередуются небольшими порциями (иными словами, работу по перемешиванию лексики как бы совершил сам Л.Н. Толстой),

отсюда и хорошее соответствие фактического роста словаря и прогнозов по формуле (1).

Т а б л и ц а 3

Объем выборки	О б ъ е м   с л о в а р я		
	фактич. $\nu^*/(N)$	по ф-ле (1)	по ф-ле (2)
500	282	319	298
1000	489	545	513
2000	824	906	862
4000	1409	1380	1395
6000	1863	1893	1840
8000	2237	2262	2225
10000	2587	2587	2590 <sup>ж</sup>

Использование теоретических формул позволяет обогатить также и возможности сравнительного анализа текстов. Например, если знать лишь объем и словарь "Пиковой дамы" ( $N = 6861$ ;  $\nu^* = 1928$ ) и отрывки из "Казаков" ( $N = 10000$ ,  $\nu^* = 2582$ ), то нельзя сделать никаких заключений об относительном богатстве (насыщенности) лексики в каждом из текстов. Калининская же кривая для "Казаков" идет уверенно выше такой же кривой для "Пиковой дамы" (ср. табл. 1 и 2). Это значит, что на случайных подвыборках равного объема из обоих текстов больший словарь будет на подвыборках из "Казаков" и, следовательно, некоторое среднее лексическое многообразие, которое принято называть относительным богатством словаря, лексической насыщенностью, лексической концентрацией и т.д., в "Казаках" выше, чем в "Пиковой даме". Полученный таким способом вывод даже надежнее, чем, например, непосредственное сравнение равных связанных подвыборок из обоих текстов, поскольку (в нашем, например, случае), на объеме  $N = 2000$  фактический словарь "Пиковой дамы" (787 слов) выше, чем словарь "Казаков" (732 слова). То же заключение о большей лексической насыщенности в "Казаках" можно сделать и из сравнения теоретических прогнозов по формуле (2) для обоих текстов (при работе с формулой (2) условной мерой лексической концентрации может служить значение параметра  $Z$ . Чем больше  $Z$ , тем большее число разных слов приходится на

случайную выборку фиксированной длины; подробнее см. Орлов, 1976; 1977.

Приведенные примеры демонстрируют те возможности лингвостатистического анализа, которые до сих пор не использовались, хотя работы В.М. Калинина были опубликованы более десяти лет назад.

Авторы признательны Р.Я. Читашвили (которым получена формула (1а)) и М.Г. Борода за участие в обсуждении смысла и возможностей применения формул (1) и (2).

### Л и т е р а т у р а

Алексеев П.М. Статистическая лексикография (типология, составление и применение частотных словарей). Ленинград, 1975.  
Андреев Н.Д. Распределительный словарь и семантические поля. - Статистико-комбинаторное моделирование языков. Ленинград, 1965.

Бектаев К.Б., Лукьяненко К.Ф. О законах распределения единиц письменной речи. - Статистика речи и автоматический анализ текста. Ленинград, "Наука", 1971.

Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М., "Наука", 1965.

Борода М.Г., Надарейшвили И.Ш., Орлов Ю.К., Читашвили Р.Я. О характере распределения информационных единиц малой частоты. - "Семиотика и информатика" (в печати).

Калинин В.М. Некоторые статистические законы математической лингвистики. - "Проблемы кибернетики", вып. II, М., "Наука", 1964.

Каширина М.Е. О типах распределения лексических единиц в тексте. - Статистика речи и автоматический анализ текста. Ленинград, "Наука", 1974.

Нешиной В.В. Оценка лексической близости текстов. В кн.: Дедушкин А.Н., Нешиной В.В., Цвиметидзе С.В. Разработка автоматизированных систем научно-технической информации. Минск, 1976.

Надарейшвили Г.Ш., Надарейшвили И.Ш., Орлов Ю.К. Некоторые статистические характеристики механизмов порождения текста. Материалы семинара "Актуальные проблемы количественной лингвистики", МГУ, октябрь 1975 (в печати).

Надарейшвили И.Ш., Орлов Ю.К. Рост лексики как функция длины текста. Сообщения АН СССР, 64, № 3, 1971.

Орлов Ю.К. Обобщенный закон Ципфа-Мандельброта и частотные структуры информационных единиц различных уровней. - Вычислительная лингвистика. М., "Наука", 1976.

Орлов Ю.К. Модель частотной структуры лексики. - Исследования в области вычислительной лингвистики. Вып. II, ч. I, изд-во МГУ (в печати, план 1977 г.).

Токарев В.П., Якубайтис Т.А. Математико-статистическая модель распределения лексем. - Математические методы в языкознании. Рига, "Зинатне", 1968.

Тулдава Ю.А. Статистический метод сравнения лексического состава двух текстов. - "Linguistica", IV, Tartu, 1971.

#### A METHOD FOR COMPLETE FIXATION OF THE TEXT IN LINGUOSTATISTICAL ANALYSIS

I. Nadareishvili and Yu. Orlov

#### S u m m a r y

A method is described for the compilation of a frequency dictionary containing full information about a text and permitting (like an index) both reverse transformation into a text and investigation of vocabulary growth within the text.

According to this method the text is transcribed on numbered cards. Thus each word use is fixed to the number of its position in the text. The fixation of the position numbers of the occurrence of each word (in increasing order) in the alphabetic list renders such a dictionary similar to an index. The arrangement of the card file according to the increasing number of the first occurrence of each word (i.e. the words in such a dictionary are presented in the order in which they first appear in the text) makes it possible to study the process of the appearance of new words (i.e. such as have not occurred earlier in the given text) with an accuracy of to within one word. The analogical ordering of words of the same frequency rank (i.e. those occurring once, twice, etc.) in the frequency dictionary (arranged according to decreasing frequency) enables one to examine the process of the appearance of new words of any given frequency.

The Supplements contain an example of the organization of a dictionary arranged in the order of the occurrence of words in A. S. Pushkin's "The Queen of Spades" and a statistical analysis of the growth curves of the vocabulary.

## ZUR IDEOLOGIEGEBUNDENHEIT DER NEUEN LEXIK IN DER DDR

### UND DER BRD

Laine Paavo

Ein in der letzten Zeit unter Linguisten Diskussionen hervorgerufenes Problem ist die Ideologiegabundenheit der Lexik. Das Wörterbuch der deutschen Gegenwartssprache von R. Klappenbach charakterisiert zwar ein Wort von mehreren Seiten, doch findet die Ideologiegabundenheit des Wortes keine exakte Kennzeichnung. Einige Hinweise über die eventuelle Ideologiegabundenheit der neuen Wörter kann die zeitliche Zuordnung (Neuwort, Neuprägung, Neubedeutung) mit dem Zusatz "DDR" bzw. "BRD" liefern. Der Zusatz "DDR" bzw. "BRD" soll unter dem Gesichtspunkt gewählt worden sein, daß "das Wort oder die Redewendung nur in einem der beiden deutschen Staaten bekannt und gebräuchlich ist" (WDG, 1974, 015). Die Einteilung der Neologismen nach ihrem Gebrauch in einem oder anderem Staat ist aber an sich problematisch. So tragen Neuprägungen Bundesrepublik Deutschland, Freie Demokratische Partei, Bundesgesetzblatt den Vermerk "BRD". Diese Bezeichnungen werden aber auch in der DDR gebraucht, wenn man über die BRD oder über die dort existierenden Sachverhalte spricht. Anders ist es mit der Neuprägung Ostzone oder mit der Neubedeutung Mitteldeutschland, die ebenfalls den Zusatz "BRD" tragen. Hier liegt der Sachverhalt außerhalb der BRD, der Gebrauch der Wörter bleibt aber auf die an die bürgerliche Ideologie gebundenen Bevölkerungskreise der BRD beschränkt.

Der Zusatz "DDR" bzw. "BRD" lenkt unsere Aufmerksamkeit auf die eventuelle Möglichkeit der Ideologiegabundenheit des neuen Wortes, doch liefert er grundsätzlich keine Informationen darüber, welche Bewertung das Wort/die Wortgruppe bzw. der Sachverhalt vom Standpunkt der entgegengesetzten Ideologie aus erfährt und ob Unterschiede im Begriffsgehalt auftreten. In manchen Fällen sind im Wörterbuch zwar zusätzliche Erklärungen eingesetzt worden, die aber den Sachverhalt nicht immer genug klar und detailliert darlegen (das soll nicht als Unzulänglichkeit des Wörterbuches aufgefaßt

werden, eher als Vielschichtigkeit und Kompliziertheit des Problems der Ideologiegebundenheit). Als Beispiel dienen hier die Wörter Arbeitgeber und Arbeitnehmer - Umschreibungen und Verhüllungen für Bourgeoisie und Arbeiter. Im Wörterbuch der deutschen Gegenwartssprache sind beide Wörter verzeichnet und mit dem Hinweis versehen: "wird auf sozialistische Verhältnisse nicht angewandt". "Dieser Hinweis trifft n.E. den Sachverhalt nicht ganz, denn wir werden auch im Hinblick auf die westdeutsche Klassengesellschaft diese Wörter nicht verwenden" (Schippan, 1968, 188).

Ideologiegebundenheit der Lexik ist in der letzten Zeit ins Blickfeld linguistischer Untersuchungen gerückt. Alle Arbeiten, die sich mit der Untersuchung der politischen Lexik befassen, haben sich mit diesem Problem auseinanderzusetzen. Anhand vieler linguistischer Arbeiten zu diesem Problem (Schippan, 1968, Fleischer, 1969, Kade, 1968) können wir schlußfolgern:

1. Politische Termini in der DDR und der BRD brauchen unterschiedliche "Übersetzung".
2. Diese unterschiedliche "Übersetzung" ist von unterschiedlichen ökonomischen und gesellschaftlichen Strukturen der Gesellschaft bedingt.
3. Die Art der ökonomischen und politischen Strukturen der Gesellschaft steht im engen Zusammenhang mit der herrschenden Klasse und ihrer Ideologie.
4. Auf der Grundlage einer Ideologie wird mit Hilfe der Sprache über die gesellschaftliche Praxis reflektiert.
5. Sprachliche Zeichen (hier: politische Termini) spiegeln auf der Grundlage unterschiedlicher Ideologien und unterschiedlicher Praxis unterschiedliches Erfassen der ideologierelevanten Realität wider und erhalten daher auch bei gleicher Form unterschiedliche Begriffsinhalte und /bzw. Wertungen.

Das Problem ist, wie weit man die Ideologiegebundenheit der Lexik erfaßt. Gehören hierher nur die Termini der "politischen Wissenschaften" (Fleischer, 1969, 475) oder gehören hierher auch die Wörter und Wortgruppen, die durch die Durchsetzung einer Ideologie veränderten oder neu entstandenen Sachverhalte, Erscheinungen, Gegenstände usw. bezeichnen?

Thea Schippan wirft ähnliche Fragen auf: "Gehören Wörter wie Liebe, Glück, Vaterland, Heimat zum politischen Wortschatz? Oder Wörter wie Erziehung, Bildung, Disziplin?" (Schippan, 1968, 177)? Sie rechtfertigt ihre Fragestellung und beantwortet die Fragen: "Schließlich sind das Wörter, deren Bedeutung durch die gesellschaftliche Stellung des Menschen zu seiner Umwelt, durch seine Weltanschauung bestimmt ist. Im weiten Sinne gehören natürlich diese Wörter zum Bereich des Politischen, des gesellschaftlichen Zusammenlebens, empfangen sie ihre Bedeutung aus der jeweiligen Klassenlage, aus der Ideologie, aus dem Wesen des Staates, der sprachregulierend eintritt" (Schippan, 1968, 177). Thea Schippan meint, daß die Ideologiegebundenheit der Lexik viel weiter zu verstehen ist als nur die Ideologiegebundenheit der politischen Termini: "Neben der Schicht der Termini der politischen Wissenschaften enthält er (politischer Wortschatz - L.P.) Wörter, die politische Tatsachen bezeichnen, an die ideologische und politische Einstellung des Sprechers gebunden sind, aber nicht definitiv festgelegt werden und somit nicht den Anforderungen an einen Terminus 'genügen' (Schippan, 1968, 182). Zu solchen politischen ideologiegebundenen Wörtern, die keine Termini darstellen, zählt Thea Schippan Wörter wie politischer Fortschritt, Friedenspolitik u.a., "Wortgut aus dem Arbeitsprozeß ..., aus den gesellschaftlichen Bereichen der Erziehung und Bildung, der Kulturpolitik, Wörter für gesellschaftliche Veranstaltungen" (Schippan, 1968, 182).

Einigen wir uns darin, daß das Zuschreiben der Ideologiegebundenheit allen Wörtern, die zur Bezeichnung der Sachverhalte, die sich durch die Durchsetzung einer Ideologie verändert haben oder die durch die Durchsetzung einer Ideologie entstanden sind, ein zu formales Herangehen an das Problem der Ideologiegebundenheit darstellt. In der deutschen Sprache läßt sich der Vergleich der verwendeten politischen Wörter samt ihrer Form und ihrem Inhalt, ihrem "kommunikativen Effekt, der durch die Verwendung" dieser Wörter "in der Rede erzielt" (Schmidt, 1969, 261) werden, gut durchführen, existieren ja Staaten unterschiedlicher Gesellschaftsordnung aber mit gleicher Sprache nebeneinander.

Wir können von den ideologiegebundenen Wörtern sprechen,



wenn unterschiedliche Bezeichnungen für einen Sachverhalt bzw. unterschiedliche Begriffsgehalte bei einer Bezeichnung bzw. unterschiedliche Wertungen sowohl bei gleicher als auch bei unterschiedlicher Bezeichnung vorliegen. Hier haben wir mit dem unterschiedlichen gedanklichen (und sprachlichen) Erfassen eines Sachverhalts bzw. unterschiedlichen Bedeutungen ein und desselben Wortes zu tun, die vom ideologischen Standpunkt des Sprechers abhängen und seine Stellungnahme den Erscheinungen der objektiven Realität gegenüber, seinen Klassenstandpunkt offenbaren.

Durch den Vergleich der politischen Wörter gelangen wir zuerst zur Form, wie die Ideologiegebundenheit auftreten kann und gewinnen Einsichten in die Bedeutungsunterschiede. Weiter müssen schon sprachliche Möglichkeiten, Mittel und ihr Einsatz untersucht werden, die die Offenbarung des ideologischen Standpunktes zulassen bzw. das Bewußtsein der Menschen in gewünschte Richtung lenken und steuern. W. Schmidt findet vier verschiedene Formen, wie die Ideologiegebundenheit des politischen Wortschatzes heutzutage auftreten kann:

1. "gleiche Bezeichnungen" für "unterschiedliche Sachverhalte" (Schmidt, 1972, 32), wobei der gleichen "Bezeichnung für unterschiedliche Begriffsgehalte eine gleiche Wertung anhaftet" (Schmidt, 1972, 32);
2. "gleiche Bezeichnung" für denselben Sachverhalt mit "unterschiedlicher Wertungskomponente" (Schmidt, 1972, 32);
3. "unterschiedliche Bezeichnungen" für denselben "Gegenstand bzw. Sachverhalt, die sich... begrifflich und nach ihrer Wertung voneinander unterscheiden" (Schmidt, 1972, 32);
4. "ideologiegebunden sind ferner alle Bezeichnungen, die mit den beiden entgegengesetzten politischen Strukturen der sozialistischen und der bürgerlichen Gesellschaft wesentlich verbunden und jeweils von deren Positionen aus benannt sind" (Schmidt, 1972, 32).

Vom besonderen Interesse sind die 2. und die 3. Form der Ideologiegebundenheit. Hier ist der Einfluß einer bestimmter Ideologie auf die semantische Widerspiegelung eines Sachverhalts ablesbar: begriffliche Gehalte decken sich teilweise (der Widerspiegelung zugrunde liegt doch ein und

derselbe Sachverhalt), Wertungen unterscheiden sich bei den Vertretern der entgegengesetzten Ideologien völlig.

Bei der vierten Form der Ideologiegebundenheit wird nicht mehr angegeben, ob diese Bezeichnungen von der Position der entgegengesetzten Ideologie gleich oder gegensätzlich bewertet werden, ob dieselben Bezeichnungen auch von der anderen ideologischen Seite verwendet werden oder ob sie eine ihren Interessen besser dienende Bezeichnung findet und ob das Finden einer anderen Bezeichnung oder Umwertung überhaupt notwendig ist. Als Beispiele für die vierte Gruppe dienen Volkspolizei, sozialistischer Wettbewerb, Bundeswehr, pluralistische Gesellschaft. Aus den Untersuchungen von Thea Schippan (anhand der Lehrbücher in der BRD) geht hervor, daß z.B. für Volkspolizei (DDR) in der BRD auch andere Bezeichnungen vorliegen: "Durch Anklänge an faschistische Wortbildungen soll beim Schüler der Eindruck entstehen, es gäbe Gemeinsamkeiten unseres Staates mit dem faschistischen Deutschland. So bezeichnen sie Volkspolizei als VOPO (Anklang an Gestapo) oder Staatspolizei (Anklang an Geheime Staatspolizei)" (Schippan, 1968, 179). Wenn wir die Sache so betrachten, so stellen Volkspolizei einerseits und VOPO und Staatspolizei andererseits "unterschiedliche Bezeichnungen" für denselben "Gegenstand" bzw. "Sachverhalt" dar, "die sich... begrifflich und nach ihrer Wertung voneinander unterscheiden" (Schmidt, 1972, 32). Nach diesem Merkmal gehören diese Bezeichnungen zur Hauptgruppe 3. Ähnlich verhält es sich mit der bürgerlichen Bezeichnung pluralistische Gesellschaft. Von uns wird die Bezeichnung als Verschleierung der antagonistischen Klassengegensätze in der kapitalistischen Gesellschaft verstanden und durch die Bezeichnung antagonistische Klassengesellschaft ersetzt.

Bei der Untersuchung der Ideologiegebundenheit der "Bezeichnungen, die mit den beiden entgegengesetzten politischen Strukturen der sozialistischen und der bürgerlichen Gesellschaft wesentlich verbunden und jeweils von deren Position aus benannt sind" (Schmidt, 1972, 32), sollte man zwischen zwei Gesichtspunkten unterscheiden:

1. Bezeichnung, Wertung, semantische Widerspiegelung e i n  
u n d d e s s e l b e n S a c h v e r h a l t s

(entweder in der DDR oder in der BRD) von der entgegengesetzten ideologischen Position aus;

2. Vergleich der Bezeichnungen, Wertungen, semantischer Widerspiegelungen ähnlicher Sachverhalte in beiden deutschen Staaten. Zu 1: Nach Einteilung von W. Schmidt hätten wir hier die zweite bzw. die dritte Form der Ideologiegebundenheit der politischen Lexik. Hier können Bezeichnungen sowohl für ähnliche, in beiden Staaten existierende Sachverhalte usw., als auch für spezifische, nur der bürgerlichen bzw. der sozialistischen Gesellschaft eigene Sachverhalte usw. erfaßt und analysiert werden. Unter diesem Gesichtspunkt vergleicht man nicht die Bezeichnung Arbeits-team mit der Bezeichnung Arbeitskollektiv, sondern man überprüft, ob marxistisch-leninistisch gesinnte Bürger der DDR den entsprechenden Sachverhalt in der BRD auch Arbeits-team nennen, ob ihre Bewertung des Sachverhalts identisch mit der bürgerlichen ist, ob es Unterschiede im begrifflichen Gehalt gibt. Genauso kann man mit den Bezeichnungen freie Marktwirtschaft oder sozialistischer Wettbewerb verfahren. Zu 2: Hier vergleichen wir wertende und begrifflich-semantische Merkmale von Arbeitskollektiv und Arbeits-team und überprüfen, ob die Bezeichnungen gegeneinander austauschbar sind, oder ob sie mit der Zeit unter bestimmten Bedingungen austauschbar werden können. Hier spielt die Ideologiegebundenheit eines Wortes eine bedeutende Rolle. Theoretisch zu untersuchen wäre, auf welche Weise, wie lange und unter welchen Bedingungen ein Wort an eine bestimmte Ideologie gebunden ist.

Im folgenden bringen wir einige Beispiele dazu, wie die bürgerliche Ideologie sprachliche Mittel zur Manipulierung und zur Verschleierung der gesellschaftlichen Prozesse einsetzt, um damit vorläufig zur Absicherung der Existenz ihrer eigenen Klasse beizutragen.

Analysieren wir den bürgerlichen Terminus Entwicklungshilfe, der im Wörterbuch der deutschen Gegenwartssprache den Vermerk "Neupräg. BRD" trägt. Die intensionale Bedeutung von Entwicklungshilfe stellt Beziehungen zu den alltagsinsprachlichen Wörtern Entwicklung und Hilfe dar, die die positiv zu wertenden Sachverhalte darstellen.

Brockhaus erklärt den Sachverhalt: "Entwicklungshilfe" - "die durch internationale Wirtschaftsorganisationen und seitens der entwickelten Industriestaaten direkt durchgeführte Maßnahmen zur Beschleunigung der wirtschaftlichen und sozialen Entwicklung der in dieser Hinsicht zurückgebliebenen Länder (Entwicklungsländer) (Brockhaus, 1968). Auch die von den sozialistischen Staaten geleistete Hilfe wird von Brockhaus mit dem Terminus Entwicklungshilfe bezeichnet.

Im WDG ist Entwicklungshilfe mit dem Zusatz "BRD" versehen, die Bedeutungserklärung stimmt mit der von Brockhaus überein (inwieweit eine so allgemeine Bedeutungserklärung wie die im WDG mit der detaillierten Darstellung in Brockhaus überhaupt zu vergleichen ist). Im WDG lesen wir, daß die Entwicklungshilfe "Unterstützung eines schwachentwickelten Landes durch einen industriell entwickelten Staat" (WDG) darstellt. Diese beiden Erklärungen erfassen nach der Meinung der marxistisch gesinnten Menschen nicht die wesentlichsten Merkmale der Sachverhalte. Anhand der gegebenen Definition des Terminus im Kleinen politischen Wörterbuch tragen wir seine wesentlichsten Merkmale zusammen:

- Form des Neokolonialismus;
- von entwickelten kapitalistischen Ländern;
- werden den ökonomisch schwachentwickelten Ländern (besonders den jungen Nationalstaaten);
- finanzielle und materiell-technische Mittel;
- unter diskriminierenden Bedingungen zur Verfügung gestellt;
- zur Verschaffung der Einwirkungsmöglichkeiten auf die politische und wirtschaftliche Entwicklung dieser Länder;
- zur Erzielung der "Profite und Zinsen für Kredite, die höher sind als die eingesetzten finanziellen und materiellen Mittel" (Kleines politisches Wörterbuch, 1973, 197).

Die Bildung dieses Terminus ist kein Verstoß gegen sprachliche Norm, denn "die Sprachgemeinschaft besitzt die Freiheit, die Motive zu wählen, die für die sprachliche Einordnung am zweckmäßigsten erscheinen, die ihren kommunikativen Bedürfnissen am besten entsprechen..." (Schippan, 1972, 50). Sie ist eher ein Verstoß gegen Sprachgemein-

schaft, der gegenüber man die wesentlichsten Merkmale des Sachverhalts verschleiert. Die Diskrepanz zwischen der intensionalen und extensionalen Bedeutung wird nur von den aufgeklärten Menschen wahrgenommen, und erst jetzt kann die intensionale Bedeutung des Terminus für sie in den Hintergrund treten. Der Terminus erfährt eine "negative Emotionalisierung" (Fleischer, 1969, 480). Der negativ emotionalisierte Terminus kann in den entsprechenden Kontexten von Vertretern der marxistischen Ideologie verwendet werden. Der ideologische Standpunkt des Sprechers offenbart sich im Kontext.

Betrachten wir noch einige bürgerliche Bezeichnungen, die eine sehr breite Verwendung in der BRD gefunden haben. Auf der metonymischen Grundlage beruht die Verwendung der Himmelsrichtung OST als einer politischen Kategorie. Ost hat durch den ständigen Gebrauch im negativen Kontext eine negative Umwertung erfahren. Durch das Vermeiden des auch in kapitalistischen Staaten auf ihre Weise im positiven Sinne verstandenen Wortes sozialistisch und durch das Ersetzen dieses Wortes mit einer Bezeichnung der Himmelsrichtung versucht man die gesellschaftliche Entwicklung in diesen Ländern zu übersehen und pejorative A s s o z i a t i o -  
n e n bei den Menschen zu erwecken. So ist die Neuprägung Ostblock eine "gegen die ideologisch-politische Einheit der sozialistischen Staatengemeinschaft gerichtete Bezeichnung für die sozialistische Staatengemeinschaft." (WDG). Aus dem ähnlichen Grund entstanden Neuprägungen Ostzone und ostzo-  
nal. Sie sollten zeigen, daß die DDR kein souveräner Staat geworden ist, daß sie immer noch eine Besatzungszone geblieben ist, die zutiefst an der Aufnahme in die BRD interessiert ist. Denselben imperialistischen Interessen dient die Bezeichnung Mitteldeutschland für die DDR.

Wenn man die eigennützige Hilfe auch noch als Hilfe bezeichnen könnte, so liegt bei der Neuprägung Entwicklungshilfe doch eine Beziehung zur Realität vor. Die Bezeichnung Ost-  
zone enthält keine in der Wirklichkeit mehr existierenden Merkmale von der DDR, in dem Sinne hat sie keinen Realitätsbezug.

Die Aufnahme der diplomatischen Beziehungen zwischen der DDR

und der BRD hat dazu beigetragen, daß Ostzone und ostzonal bereits veraltete und Mitteldeutschland veraltende Bezeichnungen in Bezug auf die DDR darstellen.

Nicht im Widerspruch zur Realität steht die Bezeichnung die ehemaligen deutschen Ostgebiete "für die zur Volksrepublik Polen gehörenden, vor 1945 im Osten des imperialistischen Deutschlands gelegenen Gebiete östlich der Oder-Neiße-Grenze" (WDG). Wichtig ist, mit welcher Absicht die Tatsache, daß diese Gebiete früher Deutschland gehörten, als Benennungsmotiv für die Prägung dieser Bezeichnung ausgewählt wurde. Diese Bezeichnung soll dem Zweck dienen, in der Sprachgemeinschaft nationale und "revanchistische" (WDG) Gefühle aufflammen zu lassen. In der Sprache der sozialistisch gesinnten Menschen wird diese Bezeichnung vermieden. Die Anerkennung der Oder-Neiße-Grenze äußert sich im Gebrauch von Bezeichnungen wie "westliche Gebiete Polens" oder "polnische Gebiete jenseits der Oder-Neiße" (Befragung).

Die Sprache als das wesentlichste gesellschaftliche Kommunikationsmittel wird in dem ideologischen Kampf zwischen den antagonistischen Klassen als Waffe eingesetzt, die im Sprachsystem verankerten Möglichkeiten werden im Interesse einer Ideologie möglichst effektiv ausgenutzt. Die Prägung motivierter Wortbildungskonstruktionen zur Bezeichnung ideologierelevanter Sachverhalte ist eine der dankbarsten Möglichkeiten der Bewußtseinsbeeinflussung der Menschen. Zur Bildung der Neuprägungen, die positive Begleitvorstellungen erwecken sollen, werden Wörter mit positivem Gehalt gewählt:

die Entwicklungshilfe (BRD),

die Friedensgrenze (DDR).

Negative Begleitvorstellungen erwecken:

die Ostzone (BRD),

der Menschenhandel (DDR).

Durch Erwecken bestimmter Begleitvorstellungen kann die Bewußtseinsentwicklung der Menschen gelenkt und gesteuert werden:

-- Erwecken und Befördern der revanchistischen Gefühle:

die ehemaligen deutschen Ostgebiete (BRD),

die Ostzone (BRD),

Mitteldeutschland (BRD);

- Erziehung des Bewußtseins im Sinne des Sozialismus, in Geiste des Friedens und der Freundschaft mit allen friedliebenden Kräften:

das Friedenslager (DDR),

das Friedensforum (DDR),

das Freundschaftstreffen (DIER).

Die Auswahl der Benennungsmotive kann aber auch der Verschleierung der Realität dienen:

die Entwicklungshilfe (BRD),

der Ostblock. (BRD).

L I T E R A T U R

Fleischer, W. Terminologie und Fachsprache im Bereich der Politik. In: Wissenschaftliche Zeitschrift der Pädagogischen Hochschule Potsdam, 1969, H. 2

Kado, O., Gossing, G. Ein Deutsch - zwei Übersetzungen? In: Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig, 1968, H. 2/3

Schippan, Th. Einführung in die Semasiologie. VEB Bibliographisches Institut. Leipzig, 1972

Schippan, Th. Die Rolle der politischen und philosophischen Terminologie im Sprachgebrauch beider deutschen Staaten und ihre Beziehungen zum allgemeinen Wortschatz. In: Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig, 1968, H. 2/3

Schmidt, W. Zum Einfluß der gesellschaftlichen Entwicklung auf den Wortbestand in der DDR. In: Deutsch als Fremdsprache, 1972, H. 1

Schmidt, W. Zur Ideologiegebundenheit der politischen Lexik. In: Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, 1969, H. 3

Wörterbuch der deutschen Gegenwartssprache (WDG). Bd. 1, 1974; Bd. 2, 1968; Bd. 3, 1973; Bd. 4, 1975

Kleines politisches Wörterbuch. Dietz Verlag. Berlin, 1973

Brockhaus Enzyklopädie in 20 Bänden. Bd. 5. F.A. Brockhaus. Wiesbaden, 1968

## К ВОПРОСУ ОБ ИДЕОЛОГИЧЕСКОЙ СВЯЗАННОСТИ НОВОЙ ЛЕКСИКИ В ГДР И ФРГ

Л. Пааво

### Р е з ю м е

Одной из проблем, вызвавших в последние годы оживленные дискуссии среди лингвистов, является вопрос об идеологической связанности (направленности) лексики. В Словаре современного немецкого языка Клаппенбах-Штейница идеологическая связанность слов особо не отмечается, но тем не менее вопрос существует. Об идеологической связанности слов можно говорить тогда, когда мы имеем дело с различным мысленным (языковым) пониманием одного и того же содержания или с различными значениями одного и того же слова, которые зависят от идеологической платформы говорящего и отражают его отношение к объективной реальности, его классовую точку зрения. Образование мотивированных словесных конструкций для выражения идеологически релевантных содержаний является одним из действительных средств оказания воздействия на сознание людей. В статье приводятся примеры идеологически связанной лексики в немецком языке в ГДР и ФРГ и выявляется реакционная сущность подобной лексики в ФРГ.



## ПОЗИЦИОННЫЙ АНАЛИЗ ПОЭТИЧЕСКОЙ СТРОФЫ

А.Я. Шайкевич

Позиционным анализом мы будем называть разновидность дистрибутивно-статистического анализа текстов, при которой изучается распределение лингвистических элементов по определенным позициям в тексте. Если в качестве целого выступает предложение (или высказывание), то его позициями могут быть, например, первое слово предложения, второе слово, третье и т.д., или последнее слово, предпоследнее и т.д. Если в процессе анализа выяснится, что те или иные слова статистически часто появляются в той или иной позиции, это будет свидетельствовать о семантической структурированности выбранного целого (в данном случае — предложения или высказывания). Чем больше позиционно связанных элементов мы обнаружим, тем более структурированным следует признать участок текста, выбранный в качестве целого. Позиционный анализ обнаруживает грамматическую структуру предложения в языках с твердым порядком слов (см. Шайкевич, 1976, 360–362), однако, его применимость к большим отрезкам текста еще подлежит проверке.

Можно ожидать, что позиционный анализ даст какие-то результаты лишь в тех случаях, когда изучаемый текст (или отрезок текста) имеет устойчивую структуру. Интересно поэтому применить позиционный анализ к поэтическим текстам с четкой строфической структурой.

Удобным предметом подобного исследования является елизаветинский сонет. Материал исследования содержится в конкордансе пяти елизаветинских сонетистов: Даниеля, Дрейтона, Сидни, Спенсера, Шекспира (Donow, 1969). Общая длина текста достаточно велика (60.294 слова).

Как известно, четырнадцать строк классического елизаветинского сонета организованы рифмой в четыре строфы: три катрена (обычно с перекрестной рифмой) и рифмованное двустишие. Рифмы могут быть замкнуты в катрены (например, у Шекспира представлена схема *abab-cdcd-efef-gg*), а могут и скреплять катрены друг с другом (как у Спенсера, где схема рифм — *abab-bcbc-cdcd-ee*). Задача позиционного анализа состоит в том, чтобы обнаружить закономерности появления слов в строках сонета и, тем самым, обнаружить внутреннюю семан-

тическую структуру.

Для частей лингвистических элементов (слов) составляется таблица распределения по строкам сонета. Примером табличной записи могут служить:

	I	2	3	4	5	6	7	8	9	10	11	12	13	14	Всего
ever	2	6	6	5	6	6	5	5	-	7	6	5	2	6	81
eue	39	22	19	13	35	26	22	19	35	22	25	16	18	22	333

Как видим, эти слова распределены по строкам сонета неравномерно. Если принять условный порог  $\frac{4}{3} \bar{x}$ , то мы обнаружим превышение порога в первой и тринадцатой строках у слова ever и в строках I, 5, 9 у слова eue. Теперь можно попарно сопоставить строки сонета и определить число общих превышений (m). Результаты приведены в табл. I. Левая нижняя часть матрицы соответствует списку 144 грамматических слов, правая верхняя - списку остальных слов (381 слов).

Таблица I

	I	2	3	4	5	6	7	8	9	10	11	12	13	14
I		34	25	20	33	27	27	16	26	17	15	19	24	30
2	8		24	29	27	32	21	22	24	22	23	27	17	18
3	13	10		26	28	28	19	27	20	16	20	27	22	26
4	3	9	10		24	28	30	35	22	16	27	24	30	22
5	20	11	13	5		30	21	24	27	23	23	21	25	25
6	4	3	5	2	6		26	42	23	25	25	26	26	22
7	8	7	9	8	8	1		30	15	21	27	19	22	23
8	3	8	9	5	6	3	6		31	33	28	34	22	28
9	20	5	11	3	12	4	7	9		26	24	25	28	29
10	5	7	6	3	7	5	6	7	6		15	21	26	16
11	8	7	9	4	6	7	7	11	14	6		25	21	23
12	7	6	7	4	4	4	7	10	10	7	11		24	31
13	19	9	8	10	19	6	6	7	23	7	13	11		37
14	12	8	10	13	9	7	10	8	13	9	13	13	21	

Долю суммы общих превышений (m) к общему числу слов списка (N) будем считать показателем сходства двух строк. Результаты показаны на рис. 1 (грамматические слова) и рис. 2 (остальные слова). Список имен и глаголов не обнаружил сколько-нибудь заметной структуры, напротив, список грамматических слов доказывает глубокую структурированность сонета: одни и те же слова часто появляются в строках I, 5, 9, 13 (в меньшей степени - в строках 3 и 11). Сходны строки 13 и 14. Есть сходство у строк 4, 12 и 14. Такое распределение

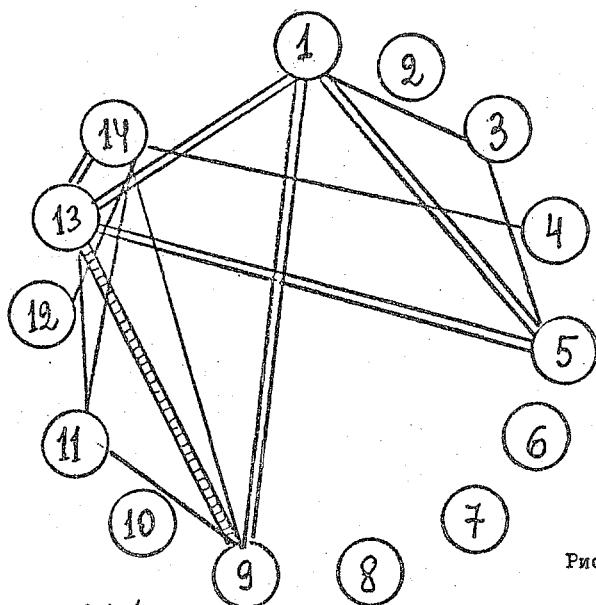


Рис. 1

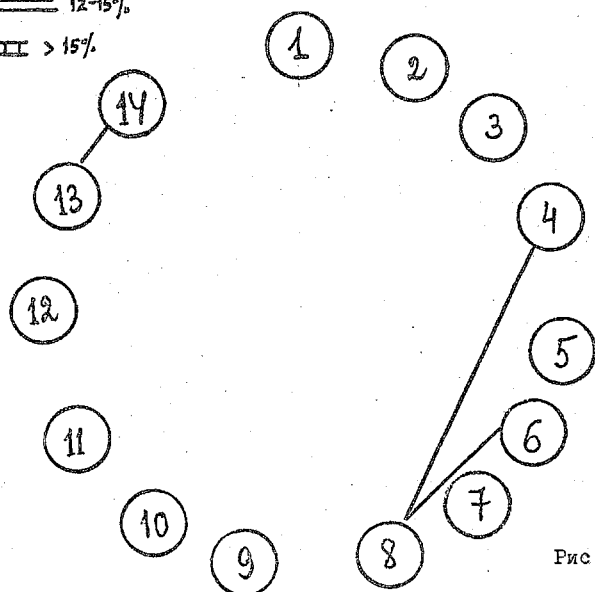
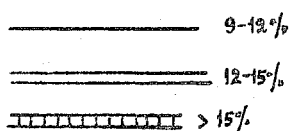


Рис. 2

строк ясно указывает на то, что внутри сонета существуют некие единства по четыре строки, внутри которых проглядывают меньшие двустрочные единства. Каждое такое единство обнаруживает внутреннюю структуру: в катренах первая строка противопоставлена остальным, в двустишиях противопоставлены нечетные и четные строки. Степень структурирования может варьировать от автора к автору. На рис. 3 показаны межстрочные связи в сонете Спенсера (по 73 грамматическим словам), на рис. 4 аналогичные связи в сонете Шекспира (по 114 грамматическим словам). Эти рисунки интерпретируются таким образом: сонет Шекспира структурирован гораздо сильнее, чем сонет Спенсера, следовательно, Шекспир реже отходит от основной грамматической схемы.

Структура двустишия носит грамматический характер: нечетные строки совпадают с началом предложения, четные — с его концом:

нечетн.                      четн.

субъектные местоимения:

I, thou, he, she, we, they                      654                      375

вопросительные слова:

how, what, when, who                      233                      118

вспомогательные глаголы:

be, do, have                      679                      449

предлоги:

by, from, of, to, upon                      661                      693

В основе структуры катрена лежит не грамматика, а сюжетный фактор. К началу катрена тяготеют следующие группы слов:

строки катрена

	I	2	3	4
I, am, me, mine, my	437	293	281	214
thou, thee, thy	123	93	72	76
Delia, Stella	30	16	10	13
ah, o, alas	82	13	19	23
cruel, fair, happy, true	125	72	70	78
love, Cupid, power, admire	208	136	101	121
see, eye, gaze, behold, hear	194	123	106	82

Итак, четверостишие обычно начинается с "расстановки сил": поэт и его любовь — вот отправное начало. Отсюда осуществляется переход к детализации внутреннего мира и к миру внешнему:

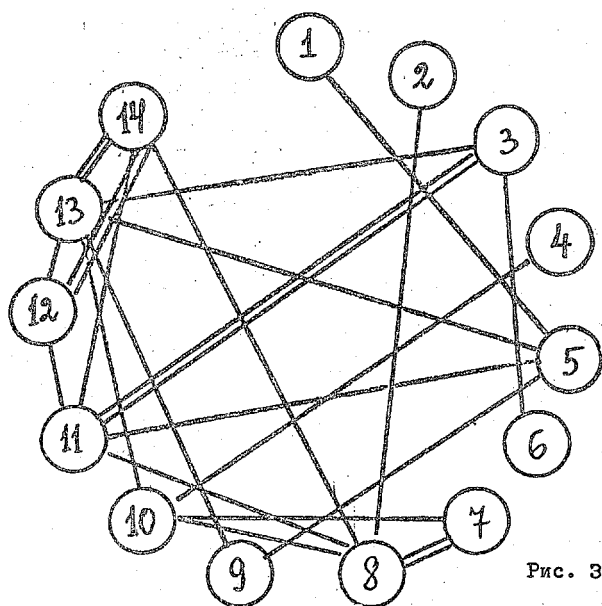


Рис. 3

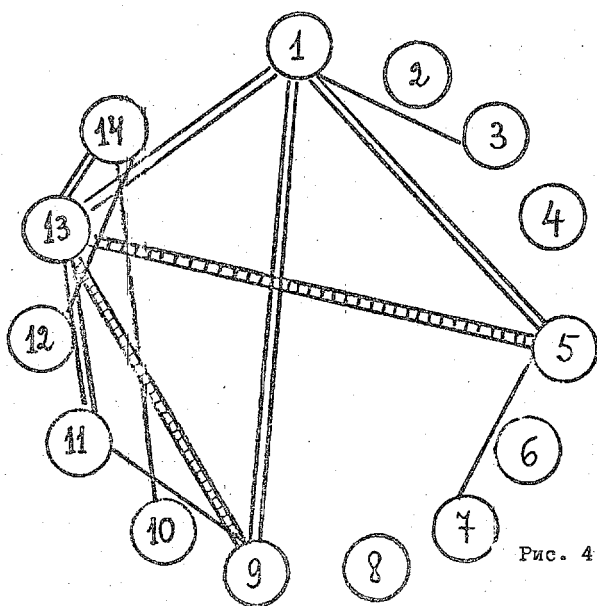


Рис. 4

строки катрена

	I	2	3	4
aspire, bear, bend, bless, bring, brought, burn, consume, decay, die, fade, get, express, grow, learn, leave, live, made, make, please, praise, prey, prove, remove, show, suffer	154	212	243	308
all, any; both, every, another, other	96	118	170	153
they, their, them	22	39	58	62
and, more, new, again	128	221	243	283
in, on, from, to	233	335	305	376

Семантическая структура катрена частично повторяется в структуре целого сонета. Об этом красноречиво свидетельствуют группы слов, представленные в табл. 2 и рис. 5.

Экспозиция сонета включает поэта и его возлюбленную, а также страсть поэта (гр. I, 2), однако, в отличие от катрена та же тема звучит и в конце сонета. Главная тема елизаветинского сонета обычно образует рамку. Как и в структуре двустишия, в начале сонета часто появляются вопросительные слова (гр. 6). Красота возлюбленной (гр. 12, 15) требует всей изобретательности поэта, чтобы найти себе достойное описание (гр. 4, 13). Альтернативная тема, начинающая сонет, — жестокость непрístupной возлюбленной (гр. 5), страдания поэта (гр. 3) в настоящем и прошлом (гр. 7-10).

От прямого названия предмета любви идет путь к орнаментальным иносказаниям, через визуальные сравнения (гр. 16), ко всему миру (гр. 17). Описание это статично (гр. 18), иногда появляется побочная тема ложности поверхностного впечатления (гр. 19).

Третий катрен вносит новый элемент динамизма в развитие сонета (гр. 20, 22, 25). Конец его предваряет заключительное двустишие, где сила страсти достигает наивысшего накала (гр. 28), преодолевая страдания прошлого (гр. 32, 33). Появляется твердая уверенность в том или ином исходе любовных страданий (гр. 27, 31, 29, 30). Снова утверждается неизменность любви (гр. I, 2, 23) уже как некая неоспоримая истина (гр. 31).

Таблица 2

	1	3	5	7	9	11	13
	2	4	6	8	10	12	14
1. I, me, mine, my, thou, thy, thee, thine, let, o	477	293	371	278	454	322	532
2. admire, adore, heart, kiss, love, lover, passion	182	106	133	112	137	108	215
3. alas, burden, captive, care, cruel, cruelty, cry, despair, fear, grief, languish, poor, prey, sad, sigh, sorrow, starve, suffer, tear, toil, torment, tyrant, unkind, weak, weary, weep, weigh, woe, work, wound	160	135	106	151	123	132	105
4. line, muse, paper, pen, verse	48	20	28	19	29	21	27
5. cruel, cruelty, disdain, frown, pride, proud, unkind	45	25	26	36	30	35	30
6. how, what, when, where, which, who, whom, whose, why	129	115	79	84	79	93	79
7. after, before, ere, last, late, long, oft, past, since, when	118	79	75	60	65	72	72
8. age, day, hour, new, old, time, young, youth	82	74	68	46	74	63	61
9. begin, begun, brought, did, found, gone, had, known, pass, saw, sought, was, were	58	50	45	42	45	39	38
10. came, come, fly, gone, run, way	39	26	25	26	32	19	23
11. brain, breast, heart, mind, soul	88	57	66	75	72	61	77
12. beauty, brow, cheek, eye, face, fair, frame, hair, hand, head, shape	193	117	174	122	157	139	122
13. as, compare, like	83	62	78	69	64	58	55

1	3	5	7	9	11	13
2	4	6	8	10	12	14

14. dwell, lay, lie, place, shore, rest, sit, stand, stay	51	47	44	50	37	43	29
15. behold, blind, eye, gaze, glass, hear, look, see, sense, sight, view	149	88	93	114	172	103	116
16. beam, black, bright, clear, color, crystal, dark, fade, hue, light, lightning, moon, ray, red, rose, shade, shadow, sun, white	58	84	83	69	59	59	50
17. air, bud, cloud, earth, field, moon, nature, rose, shade, sky, sun, tempest, tree; wing	55	52	78	58	50	50	28
18. above, at, from, in, on, out, unto, up, upon, within	138	157	183	193	162	159	135
19. conceit, deceive, false, hide, idle	7	11	14	14	21	5	6
20. great, increase, high, more, much, so, than	80	84	100	84	121	102	137
21. because, cause, effect, therefore	7	10	11	12	19	17	20
22. become, cease, change, grow, made, make, new, turn	50	61	51	44	59	80	89
23. again, back, hold, keep, remain, rest, return, same, stay, still	52	72	56	79	80	76	99
24. away, forth, from, to, up, unto	128	146	137	147	130	182	138
25. add, and, both, either, or, other, sum	98	176	126	173	142	214	177
26. he, him, his, her, she, they, their, them	83	100	90	134	115	142	143
27. must, need, shall, will	35	38	41	38	44	62	81
28. dead, death, die, eternal, grave, immortal, kill, life, live, mortal, murder	37	42	36	53	42	66	94



	<u>1</u>	<u>3</u>	<u>5</u>	<u>7</u>	<u>9</u>	<u>11</u>	<u>13</u>
	<u>2</u>	<u>4</u>	<u>6</u>	<u>8</u>	<u>10</u>	<u>12</u>	<u>14</u>
29. bad, curse, ill, worse	4	11	6	8	6	16	27
30. best, better, good, perfect, praise, virtue, well, worthy	57	57	46	68	52	71	91
31. fool, idea, know, known, learn, proof, prove, reason, seen, show, teach, think, thought, true	91	93	75	87	87	76	156
32. cure, save	1	2	-	4	4	10	11
33. bless, bliss, delight, glad, happy, triumph	28	21	19	25	22	44	45

I,  
thou  
Love  
Tears  
Muse  
Cruel  
What  
Past  
Time  
Did  
Come  
Soul  
Beauty  
Compare  
Place  
See  
Hue  
Nature  
At  
False  
More  
Cause  
Grow  
Remain  
To  
And  
He, she  
Shall  
Death  
Bad  
Good  
Know  
Cure  
Happy

Конечно, такая структура сонета отнюдь не обязательна. Очень немногие из 33 групп, представленных в табл. 2, преимущественно сконцентрированы в какой-то определенной части сонета. Таковы тема жизни и смерти (гр. 28), тема счастья (гр. 33), тема избавления от страдания (гр. 32) в конце сонета; тема лжи (гр. 19) в его середине и тема поэзии (гр. 4) - в начале. В большинстве случаев превышения над средней на тех или иных участках не слишком велики (хотя и значимы статистически).

Большие колебания обнаруживаются у отдельных авторов. Так, группа "Время" (гр. 8) крайне характерна для Шекспира; у него слова этой группы распределены по частям сонета равномерно.

Из выделенных 33 семантических групп наибольший интерес представляют те первые две группы, которые охватывают сонет. Слова, появляющиеся на периферии, как-будто задают главную тему сонета. Отсюда вытекает возможность измерить тематическую весомость слова. Поскольку из позиционного анализа следовала особая значимость строк I, I3 и I4, придадим формуле расчета следующий вид:

$$Q_x = 2 (e_i - m_i) (e_{13, 14} - m_{13, 14})$$

где  $e_i$  - доля случаев появления слова  $x$  в строке  $i$  к суммарной частоте слова  $x$ ;

$m_i$  - математическое ожидание той же доли в предположении равномерного распределения слова по строкам.

Одновременно можно определить вклад такого слова в общую тематику текста ("тематический вес"). Зададим его формулой:

$$w_x = f_x Q_x^2 10^6,$$

где  $f_x$  - относительная частота слова  $x$ .

Результаты приводятся в табл. 3.\* Как видим, формулы хорошо выделяют тему сонета: "я, ты и любовь". Слово love - неизменное тематическое слово всех авторов. Одновременно можно отметить интересные авторские вариации (см. табл. 4). Максимальные значения  $Q$  показывают следующие слова: Сидни - kiss, Stella; Даниель - Delia, love; Дрейтон - read, sea-

\* Слова, не входившие в конкорданс, отмечены звездочкой. Для таких слов подсчет производится только по Спенсеру и Шекспиру, затем результат был экстраполирован на всю выборку.

ven; Спенсер - happy, fair; Шекспир - winter, long. Вот  
два характерных примера:

Spenser, 76

Fair bosom fraught with virtues richest treasure,  
The nest of love, the lodging of delight;  
The bower of bliss, the paradise of pleasure,  
The sacred harbour of that heavenly spright.

How was I ravish'd with your lovely sight,  
And my frail thoughts too rashly led astray?  
Whiles diving deep through amorous insight,  
On the sweet spoil of beauty they did pray.

And twixt her paps like early fruit in May,  
Whose harvest seem'd to hasten now apace:  
They loosely did their wanton wings display,  
And there to rest themselves did boldly place.

Sweet thoughts I envy your so happy rest,  
Which oft I wish'd, yet never was so bless'd,

Shakespeare, 97

How like a winter hath my absence been  
From thee, the pleasure of the fleeting year!  
What freezings have I felt, what dark days seen!  
What old December's barrenness everywhere!

And yet this time remov'd was summer's time,  
The teeming autumn, big with rich increase,  
Bearing the wanton burden of the prime,  
Like widowed wombs after their lord's decease;

Yet this abundant issue seem'd to me  
But hope of orphans, and unfathered fruit;  
For summer and his pleasures wait on thee,  
And thou away, the very birds are mute;

Or, if they sing, 'tis with so dull a cheer  
That leaves look pale, dreading the winter's near.

Таблица 3

	W	Q		W	Q
I <sup>x</sup>	346	0,134	story	17	0,292
love	222	0,147	read	15	0,150
that <sup>x</sup>	215	0,105	virtue	15	0,118
you <sup>x</sup>	125	0,153	do <sup>x</sup>	15	0,051
thou <sup>x</sup>	108	0,111	day	15	0,093
my <sup>x</sup>	80	0,063	fair	14	0,060
be <sup>x</sup>	69	0,094	absence	13	0,225
shall <sup>x</sup>	69	0,137	fool	13	0,163
am <sup>x</sup>	65	0,180	good	13	0,111
have <sup>x</sup>	59	0,125	dear	12	0,095
Stella	56	0,215	deny	12	0,239
kiss	42	0,325	long	12	0,092
happy	39	0,219	mine	11	0,063
Delia	33	0,333	muse	11	0,097
unkind	33	0,277	king	9	0,200
winter	33	0,315	lack	9	0,200
since	27	0,138	ever	9	0,082
curse	24	0,321	men	9	0,104
draw	23	0,331	nothing	8	0,124
o	22	0,105	because	8	0,146
call	21	0,170	behold	8	0,107
me <sup>x</sup>	21	0,048	bid	7	0,192
fast	19	0,286	cast	7	0,178
well	18	0,112	hear	6	0,130
let	18	0,093	ah	6	0,135
dost <sup>x</sup>	17	0,100			

Таблица 4

	W	Q		W	Q		W	Q
<u>Daniel</u>			us <sup>x</sup>	79	0,092	high	50	0,200
love	454	0,212	that <sup>x</sup>	75	0,065	because	37	0,146
Delia	289	0,333	as	70	0,101	more	27	0,100
mine	94	0,176	thus <sup>x</sup>	70	0,236	joy	17	0,084
whilst	58	0,153	an <sup>x</sup>	64	0,180	sweet	13	0,056
heart	36	0,076	long	60	0,237	come	11	0,089
eye	9	0,036	say	57	0,173	sight	9	0,083
			verse	48	0,231	well	9	0,067
<u>Drayton</u>			mine	47	0,115	hope	6	0,060
love	260	0,157	be <sup>x</sup>	42	0,072			
heaven	140	0,198	not <sup>x</sup>	31	0,057	<u>Spenser</u>		
read	109	0,307	shall <sup>x</sup>	23	0,083	fair	359	0,243
sweet	40	0,096	hast <sup>x</sup>	23	0,153	I <sup>x</sup>	301	0,131
men	35	0,173	eye	22	0,063	love	194	0,170
fair	23	0,077	seem	22	0,140	happy	114	0,286
never	22	0,100	dost <sup>x</sup>	19	0,107	see	96	0,214
hope	15	0,079	are <sup>x</sup>	17	0,065	world	92	0,200
whilst	15	0,096	make	16	0,065	you <sup>x</sup>	89	0,158
fame	11	0,085	since	16	0,108	this <sup>x</sup>	75	0,150
fire	9	0,057	mind	15	0,114	shall <sup>x</sup>	70	0,125
ever	6	0,056	day	15	0,076	do <sup>x</sup>	52	0,084
o	6	0,058	art <sup>x</sup>	12	0,065	life	40	0,123
<u>Shakespeare</u>						work	28	0,146
I <sup>x</sup>	468	0,155	<u>Sidney</u>			long	26	0,107
thou <sup>x</sup>	221	0,129	Stella	267	0,215	so	20	0,044
you <sup>x</sup>	184	0,171	love	229	0,149	day	19	0,100
love	160	0,114	kiss	111	0,307	grace	14	0,108
winter	136	0,490	virtue	85	0,193	most	14	0,071
have <sup>x</sup>	81	0,136	good	55	0,168	vain	12	0,083
			true	51	0,200			

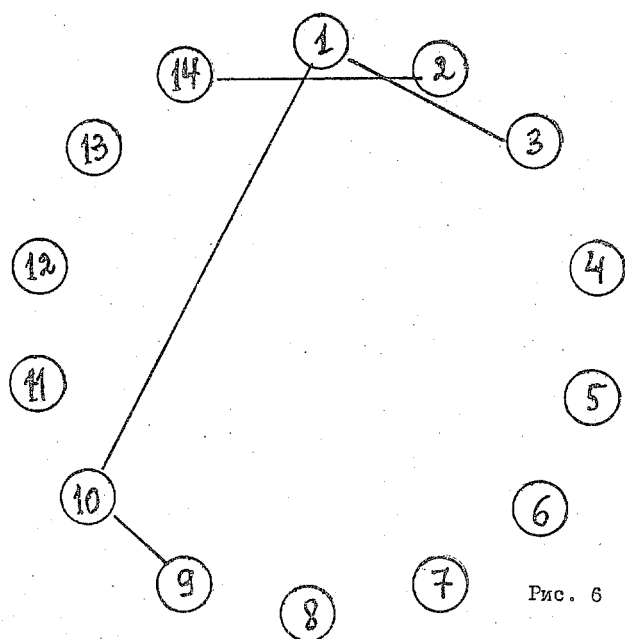


Рис. 6

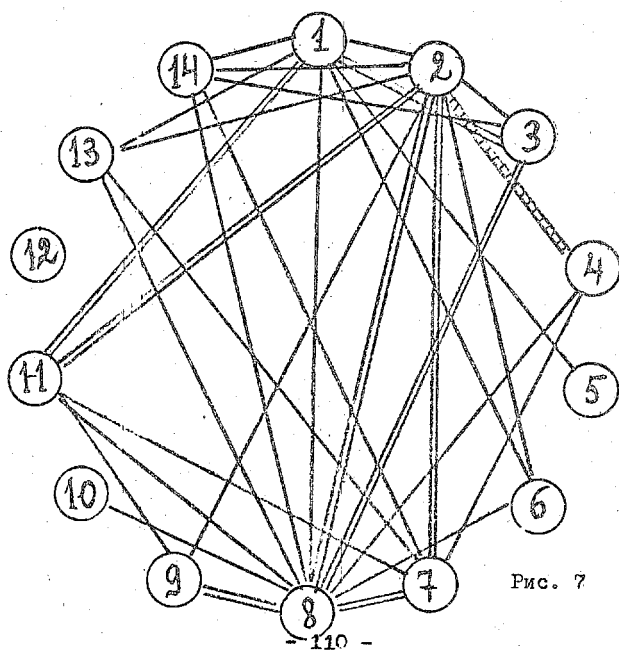


Рис. 7

Результаты позиционного анализа английского сонета интересно сравнить с аналогичными результатами для онегинской строфы. Онегинская строфа включает 14 строк (как и сонет) с жесткой схемой рифмовки: *abab-cddd-effe-gg*. Таким образом, и здесь мы сталкиваемся с тремя четверостишиями, за которыми следует заключительное двустишие. Однако в отличие от сонета, весьма автономного в семантическом отношении, строфа "Евгения Онегина" вплетена в ткань всего романа, а потому можно ожидать, что она не обнаружит такого грамматического и лексического единства, как сонет.

По списку грамматических слов (82 слова) связи между строками оказались очень слабыми (см. рис. 6). Это может свидетельствовать либо о том, что части строфы не связаны строго с предложением, либо о том, что само предложение не имеет твердого порядка слов. Список существительных, прилагательных и глаголов (176 слов) дал много слабых связей между строками, которые, однако, не укладываются в четкую структуру (рис. 7).

Четверостишие онегинской строфы обнаруживает некоторую семантическую структурированность, к его началу тяготеют такие группы слов.\*

	строки четверостишия			
	I	2	3	4
ГЕРОЙ (Евгений, Ленский, няня, Онегин, Тania, Татьяна)	126	62	56	56
НО (а, же)	120	57	74	63
КАКОЙ (где, как, когда, кто, ли, что)	170	124	133	121
ЭТОТ (сей, тот)	74	61	51	41
ОЧЕНЬ (самый, так)	33	21	19	18

Противоположную тенденцию показали отдельные слова и группы слов:

	строки четверостишия			
	I	2	3	4
И	155	187	215	235
СЕБЯ (свой)	19	35	37	39
ОЧИ (глаза, голова, грудь, лицо, ножка, рука, сердце, уста)	26	44	45	51

\* Одно из слов условно выбирается в качестве представителя всей группы и выносится вперед, остальные слова группы приводятся в скобках.



ГЕРОИ

НО

КОГДА

ВОТ

ЭТОТ

ПРОВОДЬ

БЫ

ДУМАТЬ  
КОГДА

А

МУЗА

СЕРДЦЕ

КАКИМ

И

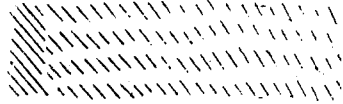
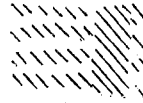
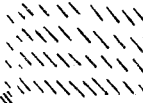
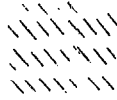
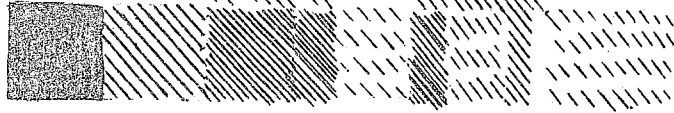


Рис. 8

Позиционные предпочтения выявлены и для строфы в целом. В начале строфы встречаются некоторые из тех же групп, которые предпочитают начало четверостишия (ГЕРОЙ, НО, ЭТОТ). Кроме них, в начале строфы преимущественно появляются группы КОГДА (тогда, уже, еще, вновь, бывало, новый, старый, пора); ВОТ (теперь, тут, здесь); ЛЮБОВЬ (любить, сердечный, чувство); ДУМАТЬ (дума, знать, мысль); БЫ (казаться, ли, мочь). По краям строфы появляются преимущественно Я (мой) и небольшая группа МОДА (модный, слава). Напротив, в середину строфы упакована группа МУЗА (перо, писать, стихи). Группы И (да, ни) и КАЖДЫЙ (везде, всегда, люди, мир, много, никто, ничто) встречаются в центре строфы и на ее конце. В конце строфы повышает свою частоту группа слов - индикаторов чувства: СЕРДЦЕ (глаза, грудь, душа, краса, нежный, ножка, огонь, очи, слеза, улыбка, уста) (см. рис. 8).

В отличие от сонета онегинская строфа не обнаруживает "ударных строк", а потому не выделяются и "тематические слова".

Основной эффект позиционного анализа - выявление внутренней структуры строфы. Эта структура оказалась весьма жесткой и определенной в случае елизаветинского сонета, напротив, она еле заметна в "Евгении Онегине". Побочный эффект позиционного анализа - выделение некоторых семантических групп слов (особенно, тематических групп в сонете).

## Л и т е р а т у р а

Шайкевич А.Я. Дистрибутивно-статистический анализ в семантике. - Сб. Принципы и методы семантических исследований. М., 1976.

Donoh, H.S. A Concordance to the Sonnet Sequences of Daniel, Dryden, Shakespeare; Sidney, Spencer. Carbondale, 1969.

## POSITIONAL ANALYSIS OF THE POETIC STROPHE

A. Shaikevitch

### S u m m a r y

Positional analysis is defined as a statistical analysis of the distribution of linguistic elements (letters, morphemes, words, etc.) in certain positions in the text. The aim of the method is twofold: to find out whether the text as a whole is structured and whether there are significant semantic (or other linguistic) characteristics of text positions. The method is applied to a poetic strophe of 14 lines with a more or less rigid pattern of rhyming:

1) Elizabethan sonnets and 2) Pushkin's "Eugene Onegin". The internal semantic structure is evident in the first case, but it is not sufficiently demonstrated in the second case. In both cases there is a strong tendency to associate certain semantic groups of words with certain positions (i.e. lines of the strophe). For the Elizabethan sonnet where the 1st, 13th and 14th lines show marked correlation it is possible to weight the thematic significance of linguistic elements (the highest values of Q in such words as love, kiss, happy, unkind, winter, etc.).

## КВАНТИТАТИВНОЕ ИССЛЕДОВАНИЕ СТРУКТУРЫ ОДНОСЛОЖНОГО СЛОВА В ЭСТОНСКОМ ЯЗЫКЕ

Ю. Тулдава

В статье исследуются закономерности относительной дистрибуции звуков в односложных словах эстонского языка. Дистрибуция звуков описывается на уровне классов — гласных и согласных. Приводятся статистические данные о встречаемости разных типов односложных слов в словаре и тексте.

Односложные слова в эстонском языке. Односложные слова составляют небольшую часть словарного запаса эстонского языка, но их роль в тексте значительна. В текстах современной эстонской художественной прозы односложные слова составляют 25-30% в авторской речи и 35-40% в речи персонажей. Изучение структуры односложного слова в эстонском языке представляет большой интерес с точки зрения дистрибуции звуков, особенно, если учесть, что подобные исследования проводились и проводятся по многим языкам мира (см., например, Vogt, 1942; Menzgerath, 1954; Seiler, 1962; Kránský, 1966; Заплаткина, 1975). Исследование и описание звуковой структуры слова, в том числе и односложного слова, имеет определенное значение при решении задач структурной типологии языков.

Известно более раннее исследование А. Рауна (Raun, 1959), который рассматривал фонотактические типы эстонских односложных слов на уровне словаря. По его подсчетам, в эстонском языке насчитывается около 3300 односложных словоформ (с учетом омонимии — 3000).<sup>ж</sup> Для сравнения можно отметить, что в родственном финском языке, в котором конечный корневой звук

<sup>ж</sup> При этом рассматриваются в качестве отдельных единиц все флективные формы слов, например, laps, last; ma, mu, mul, mind; väin, väid, väis, (ei) või, võiks. У большинства односложных слов имеется, однако, только одна такая форма (ja, ei, et, ka, siis, siin, aeg, kott, kurg, kirs, kriips и др.). Общее число лексем, имеющих хотя бы одну односложную словоформу, в эстонском языке равноприблизительно 2000. Большинство односложных слов принадлежит к наиболее древнему слою словарного запаса языка.

в большинстве случаев сохранился (например, фин. *lari* "ребенок", ср. эст. *lari*), число однословных слов совсем незначительно - около 50 (Накулинен, 1941).

Однословное слово эстонского языка можно рассматривать в качестве самостоятельного слога, состоящего в принципе из трех возможных частей: из согласного начала (*onset*), из гласного ядра (*nucleus*) и согласного конца (*coda*).<sup>12</sup> Из них обязательным является лишь ядро, состоящее из краткого гласного или (сверх) долготного гласного или дифтонга, в то время как начало и конец, состоящие из краткого или долготного согласного или сочетания согласных, могут участвовать или не участвовать в строении однословного слова.<sup>13</sup> Таким образом, наиболее общая структура эстонского однословного слова имеет следующий вид:

$$(k) \ v \ (k),$$

где *k* - согласный и *v* - гласный элемент. Возможны четыре варианта реализации общих структур, которые распределяются в словаре и в тексте следующим образом ( $\emptyset$  обозначает отсутствие элемента):

	словарь	текст
<i>k v k</i>	88,3	40,1
<i>k v \emptyset</i>	5,6	39,6
$\emptyset \ v \ k$	5,7	14,7
$\emptyset \ v \ \emptyset$	0,4	5,6
	100,0%	100,0%

Данные словаря вычислены на основе работы А. Рауна, а по тексту приводятся результаты, полученные на основе исследования однословных слов путем выборки из произведений современной эстонской художественной прозы (объем выборки - 3000 однословных словоупотреблений). Здесь и в дальнейшем все статистические подсчеты сделаны с таким условием, что *h* в начале слова считается за отдельный согласный звук (по

<sup>12</sup> В данной работе ядром считается только гласный элемент, хотя возможен и другой подход, когда к ядру причисляются сонорные согласные.

<sup>13</sup> Здесь не учитываются отдельные однословные междометия, состоящие из одних согласных, например *hm*, *pst*.

А. Рауму, звук *h* в начале слова не произносится; в данном случае сделаны соответствующие перерасчеты в его материале).<sup>12</sup>

Единицей подсчета во всех случаях выступает од н с л о ж н а я с л о в о ф о р м а (ОСФ) как на уровне словаря, так и на уровне текста, т.е. учитываются все односложные флексивные формы слов.

Рассматривая вышеприведенные данные о частности основных структур ОСФ, можно прежде всего констатировать существенную разницу в количественном распределении структур в словаре (т.е. в системе языка) и в тексте (в речи). Если в системе языка доминирует "полностью закрытый" тип *kvk* (88,3%), то в тексте его доля менее значительна (40,1%), и он имеет примерно одинаковую частотность со структурой *kvb*. Характерным для ОСФ эстонского языка является то, что в начале слова преобладает согласный элемент как в словаре (93,6%), так и в тексте (79,7%), но имеется большое различие в строении конца слова. Отношение закрытого слога к открытому в словаре составляет 16 : 1 (94% и 6%), в то время как в тексте это соотношение уравнивается - 1,2 : 1 (54,8% и 45,2%).

Представление ОСФ в виде звуковой цепи. При более подробном типологическом исследовании структуры ОСФ эстонского языка словоформы можно рассматривать в виде последовательности отдельных гласных и согласных звуков. За основу берем несколько упрощенную систему звуков эстонского литературного языка, при которой различаем лишь две ступени длительности звуков - краткую и долгую (сверхдолгую). Кроме того, не учитываются палатализация и комбинаторные варианты звуков (например, не различаются *n*, *ñ* и *ñ*). При таком подходе звуковая структура односложной словоформы эстонского языка адекватно передается в виде последовательности классов - гласных (V) и согласных (C).<sup>13</sup> Долгие звуки противопоставляются крат-

<sup>12</sup> В устной речи *h* в начале слова часто не произносится, особенно в Северной Эстонии, но все же нельзя говорить о полном исчезновении этого звука в данной позиции. В ряде случаев *h* в начале слова для многих говорящих имеет различительную функцию, например, *harv* "редкий" - *arg* "число".

<sup>13</sup> В современном литературном эстонском языке различаются следующие звуки: гласные - *a, e, i, o, u, õ* (в фонетической транскрипции), *ä, ö, ü*; согласные - сильные смич-

ким прибавлением двуеточия (v: и c:). Долгий звук считается за один, хотя он на письме (и в некоторых системах фонемной записи) передается двумя одинаковыми знаками, например, veee - три звука (cv:c), rikk - три звука (cvc:c). Эстонское односложное слово в целом считается сверхдолгим по длительности, так как в ядре или в конце слова всегда имеется хотя бы один (сверх)долгий звук. Следовательно, невозможны такие сочетания, как cv или cvc с одними краткими звуками (такие варианты могут встречаться только в связной речи). Имеются случаи, когда в односложном слове встречаются долгие звуки как в ядре, так и в конце слова, например, saak [sâk], laet [lât], <sup>\*\*\*</sup> которые при подробном фонетическом анализе следовало бы передавать через cv:c; сравни saag [sâg], laad [lâd] - cv:c. Но в данной работе мы объединяем эти разновидности в одну общую группу: cv:c. В одну группу cvcc причисляются те случаи, когда в конечном сочетании согласных один из звуков произносится долгим, например, kurg [kur̩g] и kurk [kur̩k] т.е. когда имеются разновидности cvc:c и cvcc:. При статистическом анализе встречаемости мы рассматриваем такие случаи как варианты основного типа ОСФ.

Длина ОСФ. С точки зрения типологического анализа языков имеет значение средняя длина слова и распределение слов по длине в словаре и тексте. По данным анализа встречаемости ОСФ в словаре (на основе материалов исследования А. Рауна) и в тексте (по выборке из произведений современной художественной прозы) нами установлено, что средняя длина эстонской ОСФ как единицы словаря (т.е. в данном случае практически полного инвентаря ОСФ в эстонском языке) составляет 3,72

ные (fortes) k, p, t; слабые смычные (lenes) g, b, d (в фонетической транскрипции ɢ, ʙ, ɗ); спиранты h, ʃ, v, s, ʂ, z, ʒ, r (последние четыре встречаются лишь в заимствованных словах); сонорные l, r, m, n. Некоторые звуки (t, d, s, l, n) могут быть палатализованными. Сочетания ts, tš, dz, dž (последние три в заимствованных словах) рассматриваются в соответствии с эстонской фонетикой как последовательности двух различных согласных (cc); дифтонги считаются сочетаниями двух гласных звуков (vv). В эстонском языке множество дифтонгов, наиболее частотные из них: ei, ea, ui, õi, ai, äi, au, ae, äe, õu, oi, eo, oe, ao, oe, oa.

<sup>\*\*\*</sup> В квадратных скобках дается традиционная фонетическая транскрипция эстонского языка (см. Ariste, 1953).

звука, а как единица текста - 2,74 звука.<sup>ж</sup>

Таблица I

Распределение односложных словоформ по длине  
в словаре и тексте

Длина ОСФ (количество звуков)	Доля ОСФ данной длины			
	в словаре		в тексте	
	число	%	число	%
1	4	0,1	1	0,03
2	122	3,7	1427	47,6
3	1108	33,7	963	32,1
4	1619	49,2	565	18,8
5	416	12,6	43	1,4
6	24	0,7	1	0,03
Всего	3293	100,0	3000	100,0

Сравнивая данные о встречаемости ОСФ с различной длиной в словаре и тексте (табл. I), можно констатировать, что имеются определенные расхождения в частотном распределении единиц. В словаре около половины всех ОСФ состоят из четырех звуков (49,2%). В тексте, однако, преобладают короткие ОСФ, состоящие из двух звуков (47,6% из всех односложных словоупотреблений в тексте); это - немногочисленная группа высокочастотных словоформ, составляющих лишь 3,7% общего запаса ОСФ в языке. Различия и сходства в распределении ОСФ в словаре и тексте выступают особенно рельефно в графическом изображении (рис. I). Кривые распределения ОСФ в словаре и тексте имеют разные вершины (моды), но общим для словаря и текста является то, что в обоих случаях доля 1- и 6-звуковых ОСФ является ничтожной, а 3-звуковые ОСФ покрывают одинаковую часть как в словаре, так и в тексте.

Классы и структурные типы ОСФ. Если распределить односложные словоформы эстонского языка по классам в зависимости

<sup>ж</sup> Для сравнения приводим данные о длине ОСФ, вычисленной в буквах: в словаре - 4,0; в тексте - 2,9 буквы. В русском языке средняя длина односложного слова в словаре составляет 3,6 - 3,7 буквы (Заплаткина, 1975, 13).



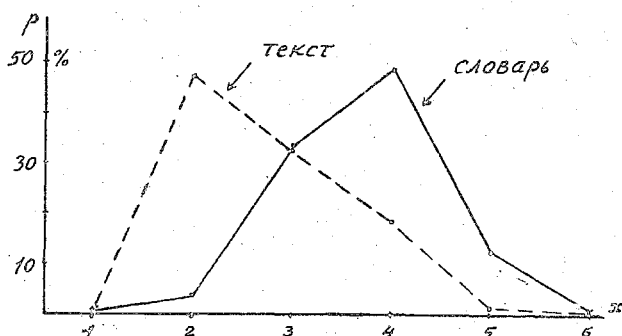


Рис. I. Односложные словоформы в словаре и тексте  
(х — длина в звуках, р — доля в процентах)

от их длины, то получим шесть классов (по количеству звуков в ОСФ), которые, в свою очередь, состоят из подклассов, или структурных типов, определяемых по сочетаниям С и V. Чем длиннее словоформа, тем больше она допускает теоретически возможных комбинаций звуков, но не все они реализуются в языке. При количественном анализе встречаемости различных типов ОСФ в словаре и тексте можно констатировать тот факт, что для каждого класса существуют два или три структурных типа, которые образуют центр данного класса, имея относительно высокую частотность появления, и остальные структурные типы, которые можно причислить к периферийным из-за малой частотности. Иногда наблюдаются существенные различия в составе центра и периферии на уровне словаря и на уровне текста. Несходство распределений структурных типов в словаре и в тексте говорит о различных закономерностях порождения односложных звуковых цепей в системе языка и их употребления в тексте, где они теснейшим образом взаимосвязаны со звуковыми цепями других порядков.

Мы уже имели возможность констатировать, что словоформы, состоящие из одного звука, в эстонском языке встречаются крайне редко. При статистическом подсчете учитывались лишь такие ОСФ, которые состоят из одного (долгого) гласного. В словаре (с общим объемом — 3293 ОСФ) насчитывает-

ся всего четыре ОСФ такого типа (öö "ночь", ee - генитив от esi "забой", междометия aa, oo), а в тексте на 3000 одно-  
сложных словоупотреблений приходится в среднем только одна  
такая форма.

Среди д в у з в у к о в ы х ОСФ, являющихся наиболее  
употребительными односложными формами в тексте, можно разли-  
чать четыре структурных типа ОСФ (табл. 2).

Таблица 2

Структурные типы двузвучковых ОСФ

Тип	Словарь число	$P_c$ (%)	Текст $P_T$ (%)	Отношение $P_T/P_c$	Примеры
CV:	60	1,8	30,5	17,0	ma, maa
V:C	34	1,0	0,7	0,7	uus, uut
VC:	19	0,6	10,8	18,0	on, õnn, ätt
VV	9	0,3	5,6	18,7	ei, au
-	122	3,7	47,6	12,9	-

Наиболее употребительным структурным типом в классе дву-  
звучковых ОСФ является CV:, однако, следует обратить внимание  
на то, что доля таких словоформ в тексте в 17 раз превосхо-  
дит их долю в словаре. Отношение частот в тексте и словаре  
( $P_T/P_c$ ) будем рассматривать как индекс итеративности (повто-  
ряемости) структурных типов в тексте. Чем больше отношение  
 $P_T/P_c$ , тем чаще повторяются в среднем словоформы данного ти-  
па в тексте. Структуру CV: имеют в эстонском языке такие вы-  
сокочастотные слова, как ja "и", ta "он, она", ma "я", me  
"мы", ka "тоже", ju "уж(е)" и др. (в связной речи они мо-  
гут произноситься и с кратким конечным гласным). Наиболее ча-  
стотными полнозначными ОСФ типа CV: являются maa "страна,  
земля", töö "работа", tee "дорога", uui "рот", kuu "ме-  
сяц", luua, puu "дерево", tee (императив от слова tegeha  
"делать" или при отрицательной форме), vee (генитив от vesi  
"вода") и др. Структурный тип CV: превосходит все другие  
типы ОСФ по частоте употребления в речи; словоформы этого  
типа составляют 30,5% от всех односложных словоупотреблений  
в тексте, но в словаре их доля незначительна (1,8%).

В классе т р е х з в у к о в ы х ОСФ можно различать 7

Структурные типы трехзвуковых ОСФ

Тип	Словарь число $p_c$ (%)		Текст $p_t$ (%)	Отношение $p_t/p_c$	Примеры
CV:C	513	15,6	10,6	0,7	lead, laat
CVC:	362	11,0	9,3	0,8	tal, tall, kott
CVV	113	3,5	9,0	2,6	kui, pea
CCV:	6	0,2	0,03	0,2	prii, truui
VCC	67	2,0	2,4	1,2	uka, ehk
VVC	43	1,3	0,4	0,3	aed, ait
V:CC	4	0,1	0,4	4,0	eest, härt
-	1108	33,6	32,1	0,95	-

Это единственный класс ОСФ, где распределения частот в словаре и тексте достаточно близкие. Лишь структурные типы CVV и V:CC имеют высокую повторяемость в тексте (см. отношение  $p_t/p_c$ ) в основном за счет сравнительно небольшого числа высокочастотных словоформ, таких как kui "если, когда", või "или", jäi "остался", hea "хороший", pea "голова", käe (генитив от käsi "рука"); eest "спереди" и др.

Следующий класс - четырехзвуковые ОСФ - самый многочисленный в эстонском языке; словоформы этого класса составляют около половины всех ОСФ в словаре (49,2%). Различается 9 разных структурных типов (табл. 4).

Таблица 4

Структурные типы четырехзвуковых ОСФ

Тип	Словарь число $p_c$ (%)		Текст $p_t$ (%)	Отношение $p_t/p_c$	Примеры
CVCC	835	25,3	8,0	0,3	kurg, kurt
CVVC	470	14,3	9,8	0,7	vaid, vait
CV:CC	160	4,9	0,7	0,15	keeld, keelt, keeks
CCV:C	76	2,3	0,2	0,09	pruun, pruut
CCVC:	55	1,7	0,03	0,02	trümm, trepp

Тип	Словарь число $p_c$ (%)		Текст $p_T$ (%)	Отношение $p_T/p_c$	Примеры
CCVV	3	0,1	0,03	0,3	kras, prao
CCCV:	1	0,03	-	-	akraa
VCCC	1	0,03	-	-	unts
VVCC	18	0,5	0,03	0,06	aust, ðalt
-	1619	49,2	18,8	0,4	-

Как видно из таблицы, все структурные типы данного класса — типичные "словарные" единицы, относительная частота которых в словаре значительно выше соответствующей относительной частоты в тексте (см. отношение  $p_T/p_c$ ). Но есть исключения, т.е. встречаются словоформы с относительно большой повторяемостью в тексте, особенно в группах CVVC (kuld "но", kõik "все", seal "там", täis "полный", käed "руки" и др.) и CVCC (nina "н", sest "ибо", just "как раз", läke "(по)-шел", miks "почему", kahe "два", kolm "три").

Следует отметить, что тип CVCC является наиболее частотным структурным типом вообще в эстонском языке среди односложных словоформ. Словоформы этого типа составляют 25,3% из всех ОСФ в словаре (835 словоформ из 3293). Выше уже говорилось о том, что в тех случаях, когда словоформа кончается на сочетание согласных, то один из них произносится обычно длиннее, чем другие, поэтому в типе CVCC объединены в сущности два подтипа: CVCC:C и CVCC:. Они распределяются в количественном отношении следующим образом:

CVCC:C — 8,3 % [kult, halb, mind; sest, miks, nahk, hetk, vaat]

CVCC: — 17,0 % [kurk, pilk, karb, kulp, vant, varg]

Пять звуковых ОСФ покрывают 12,6% словаря и 1,4% текста. В словаре зарегистрированы восемь структурных типов (табл. 5), из них наиболее частотным является тип CVVCC (7,2% в словаре и 1,2% в тексте).

К типу CVVCC относятся прилагательное laisk "ленивый", некоторые послелоготы (jaoks, realt), ряд глагольных форм условного наклонения (võiks "мог бы", peaks "должен бы", teaks "знал бы") и отдельные падежные формы элатива, аблатива и транслатива (käest "из (от) руки", peast "из головы", toast "из комнаты", neilt "от них", sealt "оттуда",

Таблица 5

## Структурные типы пятизвучковых ОСФ

Тип	Словарь число	$P_c(\%)$	Текст $P_T(\%)$	Отношение $P_T/P_c$	Примеры
CVVCC	236	7,2	1,2	0,2	loend, võiks
CVCCC	100	3,0	0,07	0,02	kimps, punkt
CV:CCC	1	0,03	-	-	raavet
CCVCC	62	1,9	0,03	0,02	sport, skalp
CCVVC	7	0,21	-	-	kloun, kleit
CCV:CC	5	0,15	0,03	0,2	kriips, prääks
CCCV:CC	4	0,12	0,03	0,3	stroof, skväär
CCCVCC:	1	0,03	-	-	sprott
-	416	12,6	1,4	0,11	-

heaks (от hea "хороший"). Встречается и партитив: naist (от naine "женщина, жена"), teist (от teine "другой").

Другие структурные типы пятизвучковых ОСФ встречаются в тексте редко, среди них много заимствований.

Что касается ОСФ, состоящих из шести звуков, то они относятся к группе малочастотных слов иноязычного происхождения. Они распределяются по пяти структурным типам (табл. 6).

Таблица 6

## Структурные типы шестизвучковых ОСФ

Тип	Словарь число	$P_c(\%)$	Текст $P_T(\%)$	Отношение $P_T/P_c$	Примеры
CCVCCC	15	0,45	0,03	0,07	prints, skunks
CCVVCC	3	0,09	-	-	praost, sveits
CCCVCC	2	0,06	-	-	sprint, splint
CCCV:CC	3	0,09	-	-	skraalt
CCCVVC	1	0,03	-	-	streik
-	24	0,73	0,03	0,04	-

Шестизвучковые ОСФ являются пороговыми; в современном эстонском языке нет односложных словоформ, состоящих из семи звуков (как, например, в русском языке слово всплеск). В рас-

рассмотренных 6 классах, образованных на основе длины слова, встретилось всего 34 структурных типа (1+4+7+9+8+5) с регулярной слоговой структурой (с гласным ядром). К ним можно при желании причислить еще нерегулярные образования С, СС и ССС (междометия *ptx[ψ]*, *[ä]*, *hm*, *st*, *ret* – всего 3 типа и 5 словоформ), в таком случае общее число структурных типов ОСФ достигает 37.

Если ранжировать рассмотренные структурные типы по частоте их появления в словаре и тексте, то пять наиболее частотных типов образуют следующие ряды:

в словаре (%)		в тексте (%)	
1. CVCC	25,3	1. CV:	30,5
2. CV:C	15,6	2. VC:	10,8
3. CVVC	14,3	3. CV:C	10,6
4. SVC:	11,0	4. CVVC	9,8
5. CVVCC	7,2	5. SVC:	9,3
<hr/> 73,4		<hr/> 71,0	

Интересно отметить, что в обоих случаях на долю пяти наиболее частотных типов (из 34) приходится свыше 70%. Обращает на себя внимание большое количество ОСФ с конечным сочетанием согласных –СС в эстонском словаре; такие словоформы образовались в результате отпадения конечных корневых гласных в процессе исторического развития эстонского языка. Первоначально эти слова были двусложными с интервокальной группой –СС–, например, *silm* < \**silmä* "глаз".<sup>‡</sup>

Анализ ОСФ на основе канонических форм. В типологических исследованиях структуры слова часто обращаются к так называемым каноническим формам (КФ) (см., например, Слипченко, 1974) для того, чтобы можно было на более общей основе проводить сопоставительный анализ структуры слова разных языков. В канонических формах не учитывается признак длительности (краткость и долгота) звуков, и дифтонги рассматриваются как простой гласный элемент. В таком случае,

<sup>‡</sup> Первоначально двусложными были и многие словоформы остальных упомянутых типов, например, *maad* (CV:C) < \**maata* (партицип от *maa* 'страна, земля') и *raad* (CVVC) < \**ritat* (номинатив множ. числа от *rida* 'ряд'), в последнем случае сокращение произошло в результате чередования звуков (*laadivaheldus*).

например, структурные типы CV:C и CVC: объединяются под общей формой<sup>CVC</sup>, а структурные типы CVVC и CVC получают общую форму - CVC.

При таком подходе пять наиболее частотных КФ эстонского языка покрывают 90,8% словаря и образуют следующий ряд по убывающей частоте встречаемости (для сравнения приводим данные некоторых других языков, см. Крамскү, 1966):

эстонский	-	CVC, CVCC, CV, CCVC, CVCCS
немецкий	-	CVC, CVCC, CCVC, CCVCC, CVCCS
английский	-	CVC, CCVC, CVCC, CV, CCVCC
венгерский	-	CVC, CVCC, VC
турецкий	-	CVC, CVCC, VC

Можно констатировать достаточно большую близость эстонского языка не только к венгерскому и турецкому языкам, но и немецкому и английскому. От последних эстонский язык отличается, главным образом, тем, что среди наиболее частотных структур отсутствует CCVCC. Три первых структуры совпадают полностью в венгерском и турецком языках.

В тексте наиболее частотными КФ односложных словоформ эстонского языка являются: CV, CVC, VC, CVCC, V, которые вместе взятые покрывают 86,7% текста.

Высокая концентрация небольшого числа высокочастотных форм в словаре и тексте является характерной чертой всех естественных языков.

Важной типологической характеристикой языков является тип и объем реализации различных структур КФ. В эстонском языке можно констатировать существование 15 КФ односложных словоформ (см. рис. 2).

На рисунке видно, что начиная от структур, состоящих из 5 элементов, не все варианты КФ реализуются в языке (например, нет структур CCCC V, CCCCC среди пятиэлементных КФ, но имеются структуры CCCVC, CCVCC, CVCCS). Сочетание из трех согласных является пороговым в начале или конце словоформы (но отсутствует тип CCCVCCS). Для сравнения можно привести примеры из других языков, в которых встречаются односложные словоформы, имеющие более трех согласных в начале или в конце слова, например, в русском языке - встреч, всхлип, взгляд; парств, средств; в белорусском - пстрык; монстр; в английском - sixths, twelfths; в шведском - avulats, nämnds; в немец-

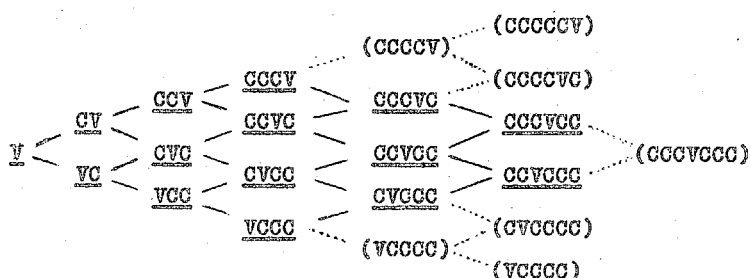


Рис. 2. Реализация канонических форм односложных слов в эстонском языке (реализованные формы подчеркнуты)

ком — schimpfst (5 конечных согласных, если сочетание pf считать за два звука).

Чтобы квантитативно выразить приведенные на рис. 2 реализации КФ в эстонском языке, можно представить систему КФ в виде таблиц сопряжения (табл. 7 и 8), где доли разных комбинаций окружения ядра выражены в процентах. Таблицы содержат различную информацию о квантитативных свойствах КФ. Например, в словаре (табл. 7) односложная словоформа наиболее часто кончается на один согласный звук (48,1%); в тексте (табл. 8), однако, преобладает нулевой согласный элемент, т.е. словоформа кончается на гласный элемент ядра (45,2%). Как в словаре, так и в тексте наиболее частотны словоформы, начинающиеся с одного согласного (86,5% и 79,3% соответственно). В словаре чаще всех встречаются слова с каноническими формами CV и CVCC; однако, в тексте, преобладают формы CV и CVC. С помощью таблиц можно решить задачу такого типа: если дано начало ОСФ, то какова будет вероятность появления конца определенной структуры (или наоборот). Можно установить, например, что если ОСФ начинается с одного согласного, то вероятность появления двух согласных в конце словоформы —  $37,4/86,5 = 0,432$  в словаре и  $29,7/79,3 = 0,375$  в тексте (индивидуальную частоту надо делить на сумму частот в соответствующей строке).

Интерес может представлять квантитативное определение степени симметричности системы КФ. Система симметрична, если



Таблица 7

Окружение ядра односложной словоформы  
в словаре эстонского языка

Число согласных		∅	В конце			∑ (%)
			1	2	3	
	∅	0,4	2,9	2,7	0,03	6,0
В	1	5,3	40,8	37,4	3,0	86,5
начале	2	0,3	4,2	2,1	0,5	7,1
ОСФ	3	0,03	0,2	0,15	0	0,4
∑ (%)		6,0	48,1	42,4	3,5	100,0

Таблица 8

Окружение ядра односложной словоформы  
в эстонском тексте

Число согласных		∅	В конце			∑ (%)
			1	2	3	
	∅	5,6	11,9	2,8	0	20,3
В	1	39,5	29,7	10,0	0,06	79,3
начале	2	0,06	0,2	0,06	0,03	0,4
ОСФ	3	0	0,03	0	0	0,03
∑ (%)		45,2	41,8	12,9	0,1	100,0

Таблица 9

Окружение ядра односложного слова (лексемы)  
в словаре немецкого языка

Число согласных		∅	В конце слова			∑ (%)
			1	2	>2	
	∅	0,6	2,1	2,5	0,6	5,8
В	1	4,4	28,7	22,1	4,9	60,1
начале	2	1,9	18,1	10,2	1,6	31,8
слова >2		0,4	1,4	0,5	0,04	2,3
∑ (%)		7,3	50,3	35,3	7,1	100,0

зеркальные структуры (CV и VC, CVC и VCC и т.д.) встречаются одинаково часто. Мы видим, однако, что в этом отношении система эстонских КФ как на уровне словаря, так и на уровне текста не отвечает полностью требованиям симметричности; например, структура CVCC намного частотнее зеркальной структуры CCVC (в словаре соответствующие частоты 37,4% и 4,2%, в тексте — 10,0% и 0,2%). Симметричность системы увеличивается в некоторой степени за счет тех случаев, когда сами КФ имеют симметричную структуру (CVVC, CVC, CVCVC и т.д.). Такие симметричные структуры покрывают в общем итоге 43,3% словаря и 35,3% текста (суммируя частоты в соответствующих диагональных клетках таблиц 7 и 8).

Сравнивая данные эстонского языка с соответствующими данными немецкого языка (табл. 9),<sup>ж</sup> можно констатировать, что симметричность системы в немецком языке (в словаре) несколько выше, чем в эстонском, например, нет большой разницы между частотами зеркальных структур CVCC и CCVC (22,1% и 18,1%; ср. в эстонском языке 37,4% и 4,2%). В немецком языке достаточно близки и распределения сумм ( $\sum$ ) строк и столбцов.

Обобщающую оценку связи между двумя признаками — звуковыми структурами начала и конца односложного слова можно выразить с помощью коэффициента взаимной сопряженности, который вычисляется по следующей формуле (Митропольский, 1971, 346):

$$\Phi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{n_{ij}^2}{n_i \cdot n_j} - 1,$$

где  $n_{ij}$  — частоты в клетках таблицы сопряженности (в данном случае таблиц 7-9),  $n_i$  и  $n_j$  — суммы частот по строкам и столбцам соответственно,  $r$  — число строк,  $k$  — число столбцов. В качестве показателя силы связи между признаками удобнее всего использовать квадратный корень из  $\Phi^2$ , т.е.  $\Phi$ . По данным эстонского словаря (табл. 7)  $\Phi = 0,102$ , по немецкому

<sup>ж</sup> Таблица составлена на основе материалов исследования П. Менперата (Menzerath, 1954), причем в отличие от наших данных в немецком языке взяты за основу словарные единицы — лексем, а не словоформы. Объем немецкого словаря односложных лексем — 2225.

Таблица 10

Канонические формы односложных словоформ  
эстонского языка с дифференциацией гласного ядра  
(по данным словаря)

№ пп.	Канонич. форма	Я д р о			Всего	
		краткий гласный	долгий гласный	дифтонг	число	%
1.	V	-	4	9	13	0,4
2.	CV	-	60	113	173	5,3
3.	VC	19	34	43	96	2,9
4.	CCV	-	6	3	9	0,3
5.	CVC	362	513	470	1345	40,9
6.	VCC	67	4	18	89	2,7
7.	CCCV	-	1	-	1	0,03
8.	CCVC	55	76	7	138	4,2
9.	CVCC	835	160	236	1231	37,4
10.	VCCC	1	-	-	1	0,03
11.	CCCVCC	1	4	1	6	0,2
12.	CCVCC	62	5	3	70	2,1
13.	CVCCC	100	1	-	101	3,0
14.	CCCVCC	2	3	-	5	0,1
15.	CCVCCC	15	-	-	15	0,5
Всего		1519	871	903	3293	-
%		46,1	26,5	27,4	-	100,0

Таблица 11

Канонические формы односложных словоформ  
эстонского языка с дифференциацией гласного ядра  
(по данным текста)

№ пп.	Канонич. форма	Я д р о			Всего	
		краткий гласный	долгий гласный	дифтонг	число	%
1.	V	-	1	169	169	5,6
2.	CV	-	916	270	1186	39,5
3.	VC	323	20	13	356	11,9
4.	CCV	-	1	1	2	0,07
5.	CVC	278	317	296	891	29,7
6.	VCC	73	11	1	85	2,8
7.	CCCV	-	-	-	-	-
8.	CCVC	1	5	-	6	0,23
9.	CVCC	239	22	38	299	10,0
10.	VCCC	-	-	-	-	-
11.	CCCVCC	-	1	-	1	0,03
12.	CCVCC	1	1	-	2	0,07
13.	CVCCC	2	-	-	2	0,07
14.	CCCVCC	-	-	-	-	-
15.	CCVCCC	1	-	-	1	0,03
Всего		918	1295	787	3000	-
%		30,6	43,2	26,2	-	100,0

словарю (табл. 9)  $\bar{\phi} = 0,141$ . Интересно отметить, что в тексте (табл. 8) зависимость между началом и концом односложной словоформы оказывается сильнее, чем в словаре (по эстонскому тексту  $\bar{\phi} = 0,239$ ). Приведенные оценки  $\bar{\phi}$  могут служить количественными типологическими характеристиками при сравнении лексических подсистем односложных слов разных языков.

При всем удобстве оперирования каноническими формами в типологических исследованиях нельзя, однако, игнорировать некоторые специфические особенности языков, которые иногда затуманиваются при анализе с помощью КФ. В частности, в эстонском языке большую роль играет контраст по длительности гласных в ядре односложного слова. Если рассматривать КФ дифференцированно по признаку заполнения ядра (краткий гласный, долгий гласный, дифтонг; см. табл. IО и II), то выясняется, что в словаре доминирует ядро с кратким гласным (46,1%), причем долгий гласный и дифтонг распределены примерно одинаково (26,5% и 27,4%). В тексте, однако, преобладают ядра с долгим гласным (43,2%), затем с кратким гласным (30,6%) и дифтонгом (26,2%). Следует подчеркнуть, что при рассмотрении звуковых цепей ОСФ в тексте мы все время оперируем "языковыми" звуками, т.е. не принимаем во внимание случаев редукции или слияния звуковых цепей в связной речи. Например, некоторые словоформы с долгим гласным могут в связной речи произноситься с кратким гласным (особенно слова типа *ja*, *ta*, *ju* и др.). В нашем анализе мы исходили из принципа инвариантности — соответствия фонетических характеристик отдельных лексических единиц в словаре и тексте.

В таблицах IО и II отражены встречаемость классов ОСФ по длине словоформ, частотность канонических форм, структурных типов и различных ядер (отдельно в словаре и в тексте). На основе этих таблиц выясняется с большой отчетливостью, что имеется принципиальное различие в количественном распределении структурных типов с кратким гласным в ядре, с одной стороны, и с долгим гласным и дифтонгом, с другой. Структурные типы с долгим гласным и дифтонгом обнаруживают большое сходство в распределении типов как в словаре, так и в тексте. На этом основании можно всю систему ОСФ разбить на две подсистемы: словоформы с кратким гласным и словоформы с долгим гласным элементом в ядре (в том числе с дифтонгом). Если построить модель подсистемы в виде так называемой параллело-

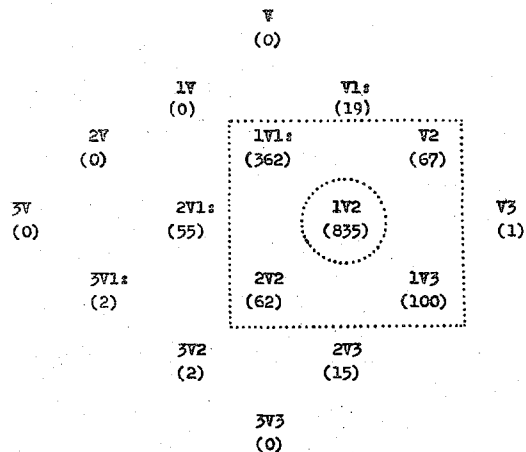


Рис. 3. Распределение типов односложных словоформ с краткими гласными в ядре (общее число в словаре - 1519)

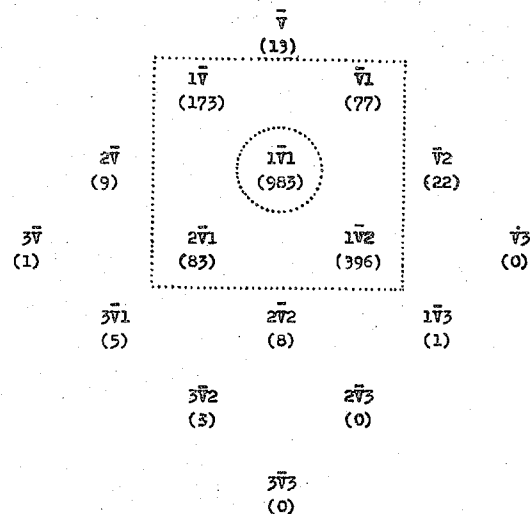


Рис. 4. Распределение типов односложных словоформ с долгими гласными или дифтонгом в ядре (общее число в словаре - 1774)

граммы Менцерата (см. рис. 3 и 4), то можно будет в обеих подсистемах выделить центр и периферию (на рисунке наиболее частотный тип, а также остальные члены центра обведены пунктиром),\* Интересно отметить, что центр в обеих подсистемах компактен и в "пространственном" смысле, т.е. образуется из наиболее частотного типа ОСФ и близких ему по структуре (близлежащих в пространстве параллелограммы) типов. При сравнении моделей особенно ясно выступает разница в размещении центров подсистем: центры покрываются лишь частично, общими типами являются *CVCV* и *CVCC* (на рисунках IV1 и IV2).<sup>жж</sup>

Заключение. В настоящей работе рассмотрены односложные словоформы эстонского языка с точки зрения их структуры и квантитативных свойств в словаре и тексте. В ходе исследования выявились некоторые системные характеристики исследуемых объектов. В частности, можно констатировать определенные закономерности дистрибуции гласных и согласных элементов в односложном слове (сопряженность начала и конца слова, характеристики ядра, связь между длиной слова и структурным типом) и выявить виды связи между отдельными элементами системы односложных словоформ (образование подсистем, выявление центра и периферии, структурная близость различных типов), а также соотношение словаря-инвентаря и функционирования элементов в тексте. Но рассматриваемая система односложных словоформ обладает целостностью лишь в относительном смысле: она является, в свою очередь, подсистемой всей совокупности словоформ языка, и задачей дальнейших исследований должно быть изучение других слоев лексики (двух-, трехсложных слов и т.д.) с целью выявления общих связей и закономерностей дистрибуции звуков. Отдельно могут быть исследованы звуковые цепи в связной речи. Наконец, интерес может представлять и изучение данной проблемы в диахроническом плане.

---

\* Центр определяется здесь как подмножество типов, имеющих наиболее высокую частотность в данном множестве типов. В настоящей работе к центру причислены пять наиболее частотных типов.

<sup>жж</sup> В параллелограмме Менцерата цифры обозначают число согласных перед или после гласного элемента, двоеточие — долготу согласного; цифры в скобках обозначают число словоформ данного типа в словаре. V — краткий гласный звук, V — долгий гласный или дифтонг.

## Л И Т Е Р А Т У Р А

- Заплаткина Н.И. Графемная структура односложных слов в славянских языках. АКД. Киев, 1975.
- Митропольский А.К. Техника статистических вычислений. М., "Наука", 1971.
- Сминченко Л.Д. Фонемная структура лексики английского языка. АКД. Киев, 1974.
- Ariste, P. Eesti keele foneetika. Tallinn, Eesti Riiklik Kirjastus, 1953.
- Hakulinen, L. Suomen kielen rakenne ja kehitys. Helsinki, 1941.
- Kránský, J. A Quantitative Analysis of Italian Mono-, Di- and Trisyllabic Words. - Travaux linguistiques de Prague, 1. Prague, Academia, 1966, 129-143.
- Menzerath, P. Die Architektur des deutschen Wortschatzes. - Phonetische Studien, Heft 3. Bonn, 1954.
- Raun, A. Monosyllabics in Estonian. - Ural-Altaische Jahrbücher, Bd. 31. Wiesbaden, 1959, 317-327.
- Seiler, H. Sprachwissenschaftliche Methoden heute: Dargestellt am Problem der deutschen Einsilber. - Studium Generale, Nr. 15, 1962.
- Vogt, H. The Structure of the Norwegian Monosyllables. - Norsk Tidsskrift for Sprogvidenskap, 12. Oslo, 1942, 5-29.

# A QUANTITATIVE INVESTIGATION OF THE STRUCTURE OF ESTONIAN MONOSYLLABICS

J. Tuldava

## S u m m a r y -

The article presents the results of a quantitative investigation of the system of Estonian monosyllabics in terms of the relative distribution of sound classes - vowels (V) and consonants (C). The author uses the material of a previous investigation of Estonian monosyllabics by A. Ravn (1959) on the vocabulary level (with some modifications) and the material of a new study of monosyllabics on the text level. In the first stage of the analysis the Estonian monosyllabics are divided into classes according to their length in sounds (see Table 1) and in each class the various types (structures) of the words and their quantitative characteristics are established both in the vocabulary and in the text (Tables 2 - 6). It is ascertained that among the 3,294 registered various monosyllabic word forms the most frequent forms on the vocabulary level are those consisting of four sounds (49.2 %) and on the text level - the forms consisting of two sounds (47.6 %). The five most frequent types in the vocabulary are CVCC, CV:C, CVVC, CVC:, CVVCC and in the text - CV:, VC:, CV:C, CVVC, CVC: (V: and C: mean a long vowel resp. consonant, VV denotes a diphthong, CC - a consonant cluster). In the second stage of the study the Estonian monosyllabics are examined in their "canonic" forms, viz. abstracting from the length of the sounds and considering the diphthong as one vowel element (V). In this case the five most frequent canonic forms in the vocabulary are CVC, CVCC, CV, CVVC, CVCCC (making up a total of 90.8 % of all forms in the vocabulary) and in the text - CV, CVC, VC, CVCC, V - (86.7 % of all forms in the text). The analysis of the inner structure of the canonic forms is carried out with the help of contingency tables (Table 7 - in the vocabulary; Table 8 - in the text) and the coefficient of contingency ( $\phi$ ) is proposed as a typological characteristic of languages. (Table 9 presents the structure of German monosyllabics on the vocabulary level.) In Tables 10 and 11 the total results of the analysis are summed up for the vocabulary and the text separately. On the whole 15 various canonic forms are registered; if differentiated according to the structure of the vocalic nucleus it appears that the nucleus with a short vowel dominates in the vocabulary (46.1 %) while the long vowel and the diphthong are distributed approximately equally (26.4 and 27.5 %). In the text, however, the long vowel is the most frequent nuclear element (43.2 %), leaving behind the short vowel (30.6 %) and the diphthong (26.2 %). According to the quantitative distribution in the vocabulary and in the text the short vowel nuclei are clearly opposed to those with long vowels or diphthongs, the last two making up another homogeneous group; when presented graphically in the form of the Menzerath parallelograms (see Figures 3 and 4) the difference in the location of the centres (five most frequent structures) and the peripheries of the groups can be clearly demonstrated.



# NAMENKUNDLICHE MISZELLANEEN

aus dem 17. Jh.

Koidu Uustalu

Die deutschsprachige Toponymie Estlands enthält zahlreiche aus dem Estnischen übernommene Namen. Die estnischen Namen sind eigentlich über das Niederdeutsche ins Hochdeutsche aufgenommen, denn das Hochdeutsche wurde erst seit dem 16./17. Jh. zur Verkehrssprache in Baltikum, die Entlehnungen stammen aber meistens aus einer früheren Zeit. Es handelt sich also um eine doppelstufige Übernahme (Grucza, 1971, 19 ff.): Estnisch → Niederdeutsch → Hochdeutsch. Im folgenden werden einige Beispiele der estnisch-niederdeutschen und niederdeutsch-hochdeutschen Namenpaare gegeben, die unter dem semantischen Aspekt gewisses Interesse bieten könnten. Die Namenbelege sind aus Thomas Hjärnes Chronik "Ehst-, Lyf- und Lettlaendische Geschichte" genommen und stammen also aus dem 17. Jh., d.h. aus der Zeit, als das Hochdeutsche sich in Est- und Livland durchgesetzt hatte, der Einfluß des Niederdeutschen aber noch ziemlich stark sein mußte.

estn. Ümera (jõgi), nd. Ymer ah

17. Jh. nach dem FluB Ymer ah (Hjärne, 83), über die Ymer ah (Hjärne, 83), an die Ymer ah (Hjärne, 85), bay Ymer ah (Hjärne, 85).

In der Chronik wird der estnische Flußname in seiner nd. Form gebraucht. Bei der Übernahme ins Niederdeutsche ist der auslautende Vokal des Namens -a in das in deutschen Gewässernamen häufig erscheinende Grundwort -a umgedeutet worden. -a geht auf lat. agua 'Wasser' (ide. \*ǵk<sup>u</sup>a 'Wasser, Fluß') zurück, das im Gotischen als Entsprechung ahwa 'Fluß', im Ahd. aha 'Fluß, Strom, Flut, Strömung', im Mhd. ahē 'Fluß, Wasser' hat und in mhd. Gewässernamen als -ach, -ache, -a seit den letzten Jahrhunderten des ersten und den ersten Jahrhunderten des zweiten Jahrtausends u.Z. belegt ist (DS II, 1970, 735, 737). Auch das Mittelniederdeutsche

kennt das Wort: a 'fließendes Wasser, Bach', belegt ebenso in der Schreibung ah (Schiller-Lübben, 1875). Demzufolge sollte die nd. Namenform eine semantische Weiterentwicklung des estnischen Namens darstellen, weil der Endvokal -a auf Grund der lautlichen Ähnlichkeit in das nd. ah selbständige Wort a umgedeutet ist.

estn. Karksi, hd. Karkus

Für den estnischen Ortsnamen Karksi ist in älteren Urkunden eine nd. Namenform belegt, in welcher der Endteil des estnischen Namens in das im deutschen Ortsnamenbestand häufige Grundwort -hausen, nd. -hus, -husen erweitert ist: 1366 Karchusen (LUB II, 1036); 1420 Carchhus (LUB IV, 3112); 1533 Carkhuis (Bfl. I, 1041). In Hjärnes Chronik kann man die übliche hd. Form des übernommenen Ortsnamens finden: Kardois (Hjärne, 384), aber die nd. Weiterentwicklung des Namens ist auch ins Hochdeutsche übersetzt: Karckhauss (Hjärne, 246, 283). Diese Übersetzung (ebenso die frühere semantische Weiterentwicklung) ist motiviert, denn Haus bedeutete früher auch 'festes Haus, Schloß' und wird von Hjärne oft in dieser Bedeutung gebraucht: das Haus Fellin (Hjärne, 84), das Haus Oberpahlen (Hjärne, 319) u.a. Bekanntlich lag in Karksi eine Burg. Der Burgname ist deutschen Ursprungs und frei, d.h. unabhängig vom estnischen Namen gebildet worden. Im Niederdeutschen hieß die Burg Soneborch. In der üblichen hd. Namenform ist das Grundwort Söhne übersetzt, der erste Bestandteil sone (auch sune) 'Sühne, endliche Ausgleichung eines Streites' wurde bei der Übersetzung in Söhne umgedeutet. In Hjärnes Chronik findet sich doch eine Namenvariante, wo die beiden Bestandteile ins Hochdeutsche übersetzt sind: Sühneburg (Hjärne, 20).

Der hd. Burgname ist deutschen Ursprungs und frei, d.h. unabhängig vom estnischen Namen gebildet worden. Im Niederdeutschen hieß die Burg Soneborch. In der üblichen hd. Namenform ist das Grundwort Söhne übersetzt, der erste Bestandteil sone (auch sune) 'Sühne, endliche Ausgleichung eines Streites' wurde bei der Übersetzung in Söhne umgedeutet. In Hjärnes Chronik findet sich doch eine Namenvariante, wo die beiden Bestandteile ins Hochdeutsche übersetzt sind: Sühneburg (Hjärne, 20).

In manchen Gegenden Estlands lebten schon in früheren Jahrhunderten Esten und Russen zusammen, und das Russische spielte bei der mehrstufigen Transposition der estnischen

Toponyme ins Hochdeutsche die Vermittlerrolle: Estnisch →  
 Russisch → Hochdeutsch, wahrscheinlich auch: Estnisch →  
 Russisch → Niederdeutsch.

estn. Jaama, hd. Jama

Eine, in Hjärnes Chronik belegte Variante des übernommenen Namens lautet Jamogorod (Hjärne, 369, 370). Sie stellt eine direkte Übernahme des in bilingualen Zuständen entstandenen "Mischnamens" dar. Da Городъ ehemals auch einen zeitweilig oder dauernd befestigten Ort, eine Verschanzung oder Festung bedeutete, ist diese Namenvariante semantisch gerechtfertigt. Übrigens, es war in jener Zeit auch eine andere, direkt aus dem Estnischen übernommene Form des Burgnamens gebräuchlich: 17. Jh. vor Jama, so ein Schloss drey Meilen vor Narva (Kelch, 335).

#### ABKÜRZUNGEN u. QUELLENMATERIALIEN

Ahd.	Althochdeutsch
Bfl. I	Est- und Livländische Brieflade. Eine Sammlung von Urkunden. I. Theil, Reval 1856-57.
Hjärne	Thomaz Hiarn's Est-, Lyf- und Lettlaendische Geschichte. Nach der Originalhandschrift herausgegeben von Dr. C.E. Napiersky. Riga/Dorpat/Leipzig, 1835.
hd.	hochdeutsch
ide.	indoeuropäisch
Jh.	Jahrhundert
Kelch	Kelch, Chr., Liefländische Historia, Dorpat, 1875.
lat.	lateinisch
LUB II, IV	Liv-, Esth- und Kurländisches Urkundenbuch nebst Regesten. B. II, IV. Hrsg. v. F.C. Bunge. Reval/Riga, 1853.
Mhd.	Mittelhochdeutsch
nd.	niederdeutsch
nhd.	neuhochdeutsch

## L I T E R A T U R

Grucza, Fr. Grundlagen der slawisch-deutschen Toponomastik im polabisch-pomoranischen Raum. - In: *Onoma* Vol. XVI (1971), 1 - 2.

*Kleine Enzyklopädie. Die deutsche Sprache. B. II.* Leipzig, 1970 (DSII).

Schiller, K., Iffken, A. Mittelniederdeutsches Wörterbuch. Bremen, 1875 - 1881.

Wasserzieher, E. Bilder aus der deutschen Sprache. 75 Aufsätze zur Sprachgeschichte. Berlin/Bonn, 1934.

Павловский, Л. Русско-немецкий словарь. Том I. Пейпциг, 1960.

### ИЗ НАБЛЮДЕНИЙ НАД ОНОМАСТИКОЙ 17 ВЕКА

К. Уусталу

Р е з ю м е

Займствование эстонских топонимических названий, ранее известных и принятых в немецком языке, происходит двухступенчато: из эстонского в средненижнегерманский; затем, после оттеснения нижнегерманского языка верхнегерманским, из нижнегерманского в верхнегерманский. В течение этого процесса произошли изменения в формах названий. Менее заметны эти изменения в семантическом аспекте: в переводе они полностью отсутствуют, однако наблюдается растирание частей названий как при переходе из эстонского в нижнегерманский, так и из нижнегерманского в верхнегерманский. В работе приводятся некоторые примеры названий из хроники Т. Ярне, т.е. названия 17 века, когда нижнегерманский язык уже был оттеснен верхнегерманским, его влияние, однако, было еще весьма заметно. В работе рассматриваются соответствия *Karchus*, *Karchhaus*; *Ümera*, *Ümerah*. Перевод был возможен при таком же переводе (из нижнегерманского в верхнегерманский) и в названиях немецкого происхождения: *Soneborch*, *Sühneburg*. На территории с эстонско-русским населением промежуточной ступенью мог быть и русский язык: *Jaama*, *Jamogorod*. В последнем случае название заимствовано немецким языком без изменений.

# THE NON-UNIQUENESS OF PHONOLOGICAL RECONSTRUCTIONS

Tiit-Rein Viitso

When speaking of protolanguages one must distinguish between two concepts of protolanguage. Following the distinction of model and theory as proposed by Reedik Palm (Hamm, 1969), I regard a MODEL as a necessary description of a set of objects and a THEORY as a sufficient description of a set of objects. A protolanguage is a model when induced from a set of dialects or languages. A protolanguage is a theory when one traces the descent and development of dialects or languages from it. The ultimate goal of historical linguistics, for any set of related dialects or languages, is not postulating a protolanguage from which all related units of the dialects or languages can be deduced, but obtaining a necessary and sufficient description of the dialects or languages. Nevertheless, a protolanguage, in order to become a theory, must first be a model.

Apparently a protolanguage as a theory can be used for improving the intermediate protolanguages, namely for selecting one of several competing solutions and even for filling certain lacunas resulting from late borrowing. But a protolanguage as a theory must not be used for archaization of an intermediate protolanguage. If, for instance, a model protolanguage has the diphthongs <sup>1</sup>ie, <sup>2</sup>uo, <sup>3</sup>oa, <sup>4</sup>ea, they must not be replaced by the monophthongs, say <sup>1</sup>e, <sup>2</sup>o, <sup>3</sup>a, <sup>4</sup>a simply because the diphthongs, theoretically, may proceed from monophthongs even when the correctness of such speculations can be evidenced by means of external data, i.e. by means of data from languages not belonging to the set whose protolanguage has the diphthongs in question. This is exactly the case with Proto-Lappish whose diphthongs have been replaced by monophthongs because of external evidence from Finnic languages. Still, postulating such theoretical intermediate stages is far from senseless. They must only not serve as a reason for eliminating the "less archaic" model protolanguages. The two types of intermediate protolanguages can be terminologically distinguished by retaining the term PROTOLANGUAGE for a model protolanguage and

calling a theoretically deduced intermediate protolanguage a PRELANGUAGE. This distinction is of especial importance here as below only problems of model protolanguages will be discussed.

The concept of the protolanguage as a model results, first, in the conclusion that a common protolanguage of a set of markedly different languages or dialects can be correctly reconstructed only on the basis of a necessary number of intermediate protolanguages.

Second, a protolanguage can be correctly induced only when the data of different object languages or dialects are unified. Any unification of the data can take place only on the basis of unified phonological descriptions of the languages or dialects. Note that existing phonological descriptions are made on the basis of very different and even unmatchable theoretical grounds, e.g. they have a very different degree of abstractness. Third, any unit of protolanguage must receive a strictly phonological shape. Mixing up units of different levels (i.e. phonetic, phonological and morpho-phonological ones) never results in greater accuracy.

Although already Yuen-Ren Chao (1954) drew attention to the presence of non-unique solutions in phonology, still one has not drawn the necessary conclusions from the phenomenon. This may be due to the belief that uniqueness/non-uniqueness depends entirely on the analyst's will or skill. However, there exist clear cases of non-uniqueness which do not depend on the analyst. One such is the following case in Estonian. Estonian has a rather complicated quantity system. In phonetic transcription besides the unmarked short sounds one has to distinguish half-long, long, overlong and extralong sounds transcribed by means of diacritics [˘ ˉ ˊ ˋ]. Note also that stops are always voiceless. Let us try to phonologize the following four partial singular paradigms:

Gloss	Nominative	Genitive	Partitive	Additive
1. tape-recorder	mák	mákki	mákki	mákki
2. Maack	mák	mákki	mákki	mákki
3. ore	mák	mági	mákki	mákki
4. magician	móg	mógi	mógi	mógi

It is meaningful to treat long, overlong and extralong vowels as sequences of two identical vocalic phonemes and to dis-

tinguish light, heavy and extraheavy accents written as /' ^ / above the first vowel of a syllable. Then we meet the problem whether (a) to contrast «single : geminate» for stops first intervocally, cf. genitive, partitive and additive, and then also word-finally or (b) to distinguish fortis and lenis stops first word-finally and then to have fortis geminates and lenis singles intervocalically:

- |     |    |       |        |        |        |
|-----|----|-------|--------|--------|--------|
| (a) | 1. | màkk  | mákki  | màkki  | mákki  |
|     | 2. | màakk | máakki | màakki | máakki |
|     | 3. | màakk | máaki  | màakki | máakki |
|     | 4. | màak  | máaki  | màaki  | máaki  |
| (b) | 1. | màk   | mákki  | màkki  | mákki  |
|     | 2. | màak  | máakki | màakki | máakki |
|     | 3. | máak  | máagi  | máakki | máakki |
|     | 4. | màag  | máagi  | màagi  | máagi  |

The first solution is historically primary as except after short first syllables (i.e. those ending in a short vowel) original final vowels have been apocopated (e.g. in the nominative) whereas the present final vowels result either from final consonant loss (e.g. in the genitive) or from contractions together with certain final losses (partitive, additive). Later borrowings have been adapted to the scheme. The second solution has given rise to degemination after long vowels, e.g. in the present author's dialect where there occurs [máki] = /máaki/ instead of /máakki/.

Non-unique solutions arise because a change makes possible the twofold solution of the new data, the original and a new one. The two solutions may coexist until a new change excludes another. In the case when the new solution gives rise to a new change one can speak about a restructuration that has triggered the change. Still there exist restructurations that occur simultaneously with the previous structururations and have neither caused nor will cause a change. Sometimes one can realize the presence of such simultaneous structururations from the linguists' (fruitless) debates about a problem, often one is simply too satisfied with what he has been taught, has read or has explained himself, to look for another solution. Nevertheless

less, the existence of non-uniqueness means, for synchronic phonology, first, that one must not be content with one feasible solution but must try to establish the whole set of permissible solutions. Second, even negative results are results: solutions whose falsehood is not trivial must be presented and their falseness proved.

When reconstructing protolanguages the situation is rather similar. As reconstructing protolanguages is not simply setting up symbols for sound correspondences and their sequences but reconstructing phonological systems and morphemes, one must not be content with establishing one possible protosystem but should try to establish the set of all possible proto-systems. Here beginning from the object languages or dialects all competing systems and subsystems whose innovational character cannot be proved must be taken into account. Otherwise one and not the best solution will be transferred from modern languages to protolanguages and so deeper on.

An example of such a possible mistake is the case of Finnic (Finnish, Estonian etc.) stem-final alternation of  $e \sim \emptyset$ . In synchrony, this alternation can be best handled in the framework of word-and-paradigm morphology establishing an alternation of stem variants. In diachrony this alternation can be handled either by means of a list of morphemes that permit the syncopation of the preceding stem-final  $e$  or by defining epenthetic or epithetic vowels in certain phonological environments. The first solution is supported by Lappish where the corresponding stems end in a vowel that is the regular correspondent of Finnic  $e$ , whereas all other Uralic languages have in that case consonantal stems. Traditionally the first solution is accepted as the main argument for reconstructing the Proto-Finnic, Proto-Finno-Ugric and Proto-Uralic  $^H e$ -stems and even for reconstructing a set of Proto-Nostratic  $^H i$ -stems (Nostratic languages include the Uralic, Dravidic, Altaic, Indo-European, Khartvelic and Afro-Asiatic families). Although I do not claim that this approach is erroneous it is evident from this example that mistakes tend to grow in the course of further reconstructing.

Another curious case of not considering all possible solutions is that of reconstructing Proto-Armenian stop and



affricate series as in Table 1. Having proved that the data in Table 1 cannot be explained by means of a split hypothesis on the basis of dialects 2, 4 or 7, it is easy to see that the Proto-Armenian series can hardly be identical with those of dialects 1 or 5 as then either unnatural rules (flip-flops) are needed or an unprovable fourth stop must be introduced. The triadic series of dialects 3 and 6 enable one to derive those of all other dialects in a similarly simple way and are thus both equally good candidates for having preserved the Proto-Armenian series.

Table 1

Correspondence Set	Dialect						
	1	2	3	4	5	6	7
A	th	th	th	th	th	th	th
B	t	t	t	d	d	d	d
C	dh	dh	dh	dh	dh	dh	dh

As the Old Armenian literary language is considered to have a series identical either with that of dialect 1 (spoken in the Ararat Valley and including the East Armenian literary language) or with that of dialect 3 (spoken on the left bank of the Euphrates in Central Armenia) even the Proto-Armenian series is considered to be identical either with that of dialect 1 or that of dialect 3: let up to now nobody has tried to counterevidence the fact that dialect 6 (spoken in Little Armenia) seems to be a still better candidate for being the most archaic in the point because of having quite regularly *d* and *dh* for (theoretical) Proto-Indo-European *d* and *dh*.

There is maybe one case when only one of two contradictory solutions can be adopted. This is the case when one cannot prove whether the data of object languages or dialects result from a split or a merger. Here it is safer to accept the merger hypothesis and to reconstruct more proto-units than may be true than to accept the split hypothesis and to blend potentially different units. A case in point occurs in the reconstruction of Proto-Athapaskan where one

may reconstruct (a) three or four stop/affricate and fricative series with (b) either four or five members each. Hence it is a question of whether to establish  $3 \times 4$ ,  $3 \times 5$  (cf. Sapir 1931, Hoijer (1960),  $4 \times 4$  (Krauss 1964) or  $4 \times 5$  proto-consonants (Viitso 1976) for the same set of correspondences. Until nothing is proved it is reasonable to have  $4 \times 5$ .

To sum up: Establishing a whole set of non-unique solutions provides us a better chance of catching a piece of historical truth.

### L i t e r a t u r e

Chao, Y.R. (1934). The non-uniqueness of phonemic solutions of phonetic systems. - Bulletin of the Institute of History and Philology, Academia Sinica 55:4, 363-97. Reprinted in: M. Joos (ed.), Readings in Linguistics, Washington 1957, 38-54.

Hoijer, H. (1963). The Athapaskan languages. - Studies in the Athapaskan Languages, Berkeley and Los Angeles (University of California Publications in Linguistics 29), 1-29.

Krauss, M.E. (1964). Proto-Athapaskan-Eyak and the problem of Na-Dene: Phonology. - International Journal of American Linguistics 30, 118-131.

Sapir, E. (1931). The concept of phonetic law as tested in primitive languages by Leonard Bloomfield. - Methods in Social Science: A Case Book, Chicago, 297-316. Reprinted in: Selected Writings of Edward Sapir, Berkeley and Los Angeles 1949, 73-83.

Viitso, T.-R. (1976). On classifying the Athapaskan languages. - Linguistica VIII, Tartu (Acta et Commentationes Universitatis Tartuensis 401), 147-163.

Палы, Р. (1969). К вопросу об основании математической лингвистики. - Труды вычислительного центра 18, Тарту (Тартуский государственный университет), 69-104.

## НЕОДНОЗНАЧНОСТЬ ФОНОЛОГИЧЕСКИХ РЕКОНСТРУКЦИЙ

Т.-Р. Вийтсо

### Р е з ю м е

Здесь праязык рассматривается как модель, т.е. необходимое описание заданного множества языков и диалектов. Тогда, во-первых, общий праязык резко отличающихся один от другого языков корректно реконструируем только на основе необходимого количества промежуточных праязыков и, во-вторых, праязык корректно реконструируем только на основе единого фонологического описания рассматриваемых языков. Из проблемы неоднозначности фонологических решений следует в синхронном плане; что следует найти не одно допустимое решение, а множество всех допустимых решений. При реконструировании праязыков для установления наилучшей прасистемы, аналогично, следует установить не одну допустимую прасистему, а множество всех допустимых прасистем, учитывая при том все допустимые начальные и промежуточные системы, чей инновационный характер не доказан.

## ОБ ОРГАНИЗАЦИИ ИНФОРМАЦИИ В СЕМАНТИЧЕСКОЙ СТРУКТУРЕ СЛОВ, ВЫРАЖАЮЩИХ ДЕЙСТВИЯ

Х. НЫМ

Проблема описания семантической структуры выражений, обозначающих целенаправленные действия людей, все больше привлекает к себе внимание исследователей как в лингвистической семантике, так и в прикладных исследованиях, моделирующих процессы понимания языка. Вряд ли найдется слово, над семантической структурой которого больше спорили бы в семантической литературе последних лет, чем английское слово *to kill* 'убивать' (McCawley, 1968; 1972; Fodor, 1970; Кас, 1972; Shibatani, 1972; Wierzbicka, 1975). Разумеется, проблема не столько в значении самого этого слова, сколько вообще слов, обозначающих целенаправленные действия. Описанию той же категории слов так или иначе посвящено и большинство работ, занимающихся анализом слов в терминах презуппозиций, импликаций и утверждений (Fillmore, 1971; Karttunen, 1971; Lakoff, 1972; Jackendoff, 1975; 1976; Почепцов, 1976 и многие другие).

Оживленный интерес к анализу слов и выражений, обозначающих действия людей, проявляют и авторы прикладных исследований, целью которых является моделирование процессов речевого общения. Центральное место занимает понятие действия в лингвистической концепции Р. Шенка (Schank, 1975), а также в работах Г. Скрагга (Scragg, 1975), И. Уилкса (Wilks, 1975). Причина этого весьма очевидна. Названные исследования являются, как правило, частями более общих проектов разработки систем искусственного интеллекта, а главная задача последних — моделировать целенаправленную деятельность разумных существ. Язык же является одним из основных источников, откуда можно черпать данные о внутренних механизмах этой деятельности: "Естественный язык используется для описания, объяснения и оценивания человеческой деятельности ... Так как естественный язык представляет собой метаязык, на котором описывается и объясняется человеческая деятельность, то он исследуется с целью выявления тех понятий, которые характерны для обсуждения этой деятельности". (Schmidt, 1975: 215).

Также и в теоретической лингвистике в последнее время

заметно возрос интерес к описанию деятельности людей: это связано с расширением круга интересов лингвистики и с постепенным переходом от описания языка как формальной системы, "кода", к комплексному исследованию языкового общения, причем это общение понимается как разновидность целенаправленной деятельности (Yngve, 1969; 1975; Звегинцев, 1973; 1976; Почешов, 1975; 1976; Бил, 1973; 1977 и др.).

Таким образом, семантическая структура лексики целенаправленной деятельности представляет интерес прежде всего как материал, исходя из которого можно конструировать модель человеческой целенаправленной деятельности. Нужно учитывать, разумеется, что это будет не какая-нибудь строго научная модель деятельности, а модель, отражающая представления здравого смысла о структуре деятельности. Но именно такая модель — как часть общей "модели мира" человека — и требуется, чтобы объяснить как люди интерпретируют действия друг друга, описывают их и понимают тексты, описывающие их.

В лингвистической семантике, а в большей мере также и в прикладных исследованиях действия описывались и описываются весьма "механически", например, анализируя слово убить как "каузировать умереть" (и умереть как "становиться неживым"), сообщать как "каузировать знать" и т.д. При таком подходе, однако, полностью теряется специфика выраженных этими словами явлений как целенаправленных действий. Действия всегда имеют две стороны, но лишь одной из них является непосредственный, внешне наблюдаемый (физический) акт деятельности. Другую же сторону образуют различные факторы, которые относятся к психике агента деятельности и образуют субъективную основу соответствующего действия: всякое действие направлено на достижение какой-то цели, его основой является решение субъекта совершить данное действие, а его исходным пунктом какое-то желание или потребность субъекта, которое или которую он старается удовлетворить через данное действие, и т.д. Именно наличие этого внутреннего, субъективного аспекта отличает действия от физических или механических процессов, происходящих в природе. И прежде всего в терминах понятий последнего типа люди интерпретируют действия друг друга и, в частности, называют их теми или другими именами.

О важности различения указанного внутреннего аспекта в структуре деятельности и о языковой реальности этого аспекта мы говорили в другой работе (Бил, Сильдмяз, в печати). Скон-

структурированная там "модель целенаправленной деятельности" (которая служит основой для описания значений слов, обозначающих действия людей) состоит из следующих структурных компонентов:

- (1) S WANT P "S желает, чтобы имело место P";
- (2) S SUPP (D CAUS P) "S полагает, что действие D приводит к P";
- (3) S CONSID D "S обдумывает, рассуждает, делать ли ему D";
- (4) S DEC D'S принимает решение делать D";
- (5) S INTEND D "S имеет намерение (интенцию) делать D";
- (6) S DO D "S делает D";

Отмечаем, что из приведенных компонентов образуется только общий "каркас" структуры деятельности. В структуре действий, выражаемых конкретными словами, могут встречаться и другие компоненты (вдобавок к конкретизации значений переменных P и D).

В качестве доказательства того положения, что при интерпретировании значений слов и выражений действительно оперируют такими понятиями, можно привести различные языковые факты. Так как сущность этих элементов не является здесь нашей основной проблемой, то остановимся только на некоторых из этих фактов.

Двумя "критическими" элементами в приведенной выше схеме являются компоненты (1) и (4), идентифицирующие, соответственно, желание (в широком смысле) субъекта, которое побуждает его к деятельности, и решение совершить действие D (через которое надеется удовлетворить свое желание). "Начальным толчком" всякой деятельности является какое-нибудь желание, потребность, влечение (добиться или избежать определенной ситуации) и ни одна целенаправленная деятельность не может произойти без сознательного решения агента.

Иметь определенную точку зрения на строение "субъективной стороны" деятельности вместе с тем означает занять определенную позицию на внутреннюю структуру самого агента, т.е. принять определенную "наивно - психологическую модель" человека - агента деятельности. Исходя из приведенных выше рассуждений мы можем говорить о двух основных "механизмах" в психике человека - агента: одним является механизм, порождающий желания, влечения и потребности, другим - механизм принятия решений и их осуществления. Первый из этих механизмов

представлен в приведенном списке элементом S WANT P; другой — элементами S DES D и S INTEND D. (Вполне очевидно, что кроме названных двух механизмов должен существовать еще по крайней мере один механизм, который связывает между собой первые два. Это — механизм знаний, предположений и оценок, представленный в приведенном списке компонентами (2) и (3); но на этом механизме мы здесь останавливаться не будем.

В случае элементов DES и INTEND субъект S выступает как сознательный агент. Механизм, компоненты которого представляют эти элементы, подчиняется сознательному управлению со стороны человека. Элемент WANT же репрезентирует механизм, который в представлениях здравого смысла трактуется не подчиняющимся воле человека. Наоборот, человек по отношению к этому механизму выступает как "пациент". Реальность указанного различия и соответствующих механизмов, а также их важность с точки зрения истолкования деятельности четко проявляется в том, как о соответствующих явлениях принято говорить.

О желаниях, потребностях, влечениях и т.п. говорят, что они "возникают", "появляются", или же "проходят", "исчезают"; им "сопротивляются" или "поддаются"; с ними "борются", их "удовлетворяют" и т.д. Ни одно из этих выражений не подходит для характеристики решений или целей. Особенно ярко выявляется своеобразная роль механизма желаний и потребностей в "модели психики" здравого смысла, если поглубже вникнуть в употребление выражений поддаваться и сопротивляться. Сопротивляться или поддаваться можно, с одной стороны, внешнему нажиму, например, уговариванию или заманиванию со стороны других людей; но, с другой стороны, сопротивляться или поддаваться можно и нажиму, исходящему "изнутри", — своим желаниям, влечениям, страстям и т.п. В первом случае, т.е. в случае "внешнего" нажима, не возникает особых трудностей при ответе на вопрос, кто является агентом (источником) нажима или воздействия и кто ему сопротивляется (или поддается). Но во втором случае ответить на данный вопрос не так-то просто. Единственное удовлетворительное решение этой проблемы мы видим в различении двух механизмов внутри одного и того же субъекта: источником воздействий, нажима является механизм желаний, а механизм принятия решений сопротивляется или же поддается этим воздействиям.

Далее, доказательство о том, что решения и намерения в

истолковании здравого смысла рассматриваются как реальные, притом "критические" компоненты деятельности, можно получить из анализа таких слов, как уговаривать, настаивать, манить, и т.п.

Цель агента ( $S_1$ ) этих действий заключается в том, чтобы заставить другого субъекта ( $S_2$ ) совершить определенное действие  $D$ . Но, как правило, уговаривание, настаивание и т.п. не продолжаются до того момента, когда  $S_2$  действительно совершит  $D$ . Критической "границей", начиная от которой указанные типы воздействия теряют смысл, скорее является решение субъекта  $S_2$  сделать  $D$ . При этом между принятием решения  $S_2$  и самим действием  $D$  может быть какой угодно большой промежуток времени. Но если  $S_1$  продолжает оказывать воздействие на  $S_2$  и после того, как  $S_2$  решил совершить  $D$  (и  $S_1$  это известно и он верит в искренность решения  $S_2$ ), то было бы неестественно называть деятельность  $S_1$ , например, уговариванием. Особенно показательным в этом отношении является слово переубеждать: переубеждать можно кого-нибудь именно тогда, когда он уже решил сделать что-то, но еще не реализовал это решение.

Теперь уместно задать вопрос, как описанные выше компоненты организованы в значениях слов, выражающих целенаправленные действия. Обычно в работах, в которых вообще рассматриваются такие "психологические" компоненты деятельности, этот вопрос не затрагивается. Например, в теории речевых актов, где весьма детально различают всякие предположения, намерения и пр. субъектов соответствующих актов, эти компоненты, как правило, просто перечисляют. Таким же образом обычно поступают при "презуппозиционном" анализе слов (Fillmore, 1971 и др.). Но предпосылкой такого перечисления является тот факт, что соответствующие компоненты-утверждения были бы действительно все вместе, в одно и то же время. Поэтому, если никакие специальные отношения между этими компонентами не указываются, то можно сделать заключение, что семантическая структура соответствующего слова представляет собой конъюнкцию перечисленных компонентов. Но является ли такой подход эмпирически оправданным?

Для наглядности рассмотрим конкретный пример. Пусть это будет структура слова просить:  $S_1$  просит  $S_2$ , чтобы  $S_2$  сделал  $D$ . Так как нас интересует только организация материала внутри семантической структуры этого слова, то точность



самого описания значения здесь не является существенной (о более детальном анализе семантики слова просить, хотя с несколько иных позиций, см. Чумаков, в печати).

Значение слова просить мы можем описать следующим образом:  $S_1$  желает, чтобы  $S_2$  сделал  $D$ ; он предполагает, что  $S_2$  сам не собирается делать  $D$ , но что если он ( $S_1$ ) сообщает ему о своем желании, то  $S_2$  будет делать  $D$ ; исходя из этих предположений,  $S_1$  сперва обдумывает, сообщить ли  $S_2$  о своем желании, затем приходит к решению сообщить, и через соответствующее намерение (интенцию) это решение приводит  $S_1$  к тому, что он сообщает  $S_2$ , что он желает, чтобы  $S_2$  сделал  $D$ . Используя введенные выше обозначения, мы можем представить этот материал в более эксплицитной форме:

- (1)  $S_1$  WANT, чтобы сделал  $D$ ;
- (2)  $S_1$  SUPP (а) что  $S_2$  сам не собирается делать  $D$ ;
- (3) (в) если он ( $S_1$ ) сообщает  $S_2$  о своем желании, то  $S_2$  будет делать  $D$ ;
- (4)  $S_1$  CONSID, сообщить ли ему  $S_2$ , что он ( $S_1$ ) хочет, чтобы  $S_2$  сделал  $D$ ;
- (5)  $S_1$  DEC сообщить  $S_2$ , что он хочет, чтобы  $S_2$  сделал  $D$ ;
- (6)  $S_1$  INTEND сообщить  $S_2$ , что он хочет, чтобы  $S_2$  сделал  $D$ ;
- (7)  $S_1$  INFORM (= сообщает)  $S_2$ , что он хочет, чтобы  $S_2$  сделал  $D$ .

При таком представлении интуитивно совершенно ясно, что было бы неправильно интерпретировать компоненты (1) - (6) просто как (конъюнктивное) множество логических условий, в контексте которых действие, представленное компонентом (7) - сообщение о своем желании, оказывается просьбой. Перечисленные компоненты упорядочены определенным образом. Компоненты (1) - (6) представляют как будто "предысторию" рассматриваемого действия в виде тех процессов, которые привели  $S_1$  к данному действию.

Эта история представляет собой упорядоченную во времени и цепь событий. Далеко не безразлично, в каком порядке мы перечисляем компоненты (1) - (7). Только по отношению к первым трем компонентам с элементами WANT и SUPP это упорядочивание не имеет особого значения. Но уже этап "обдумывания", представленный компонентом CONSID, может начинаться только при условии, что соответствующее желание и предположения уже имеются у субъекта  $S_1$ . Ведь фактически это составляет  $S_1$  обдумывать, обращаться ли ему к  $S_2$ . Далее, акт

принятия решения DEC может иметь место только после этапа обдумывания (CONSID). Таким же образом INTEND является результатом принятия решения и, следовательно, может совершаться только после DEC.

При более подробном рассмотрении можно обнаружить, что временные соотношения между компонентами являются еще более сложными: мы можем не только упорядочить их по их "начальным точкам", а также и сказать, когда один или другой этап кончается.

Желание и предположения, представляемые первыми тремя компонентами, очевидно, должны сохраняться в течение всей "истории" действия, а также и во время самого действия. Далее, этап обдумывания CONSID может длиться только до принятия решения (DEC): обдумывание вопроса - делать ли  $p$  - кончается в тот момент, когда принимается решение (делать  $p$  или не делать). Принятие решения DEC является не процессом, а актом. Естественно, что он не может продолжаться во время следующего этапа (INTEND), который является его результатом. Сам INTEND, однако, должен длиться не до начала следующего этапа - самого действия, а также и во время действия, ведь именно INTEND поддерживает и управляет процессом совершения действия.

Таким образом, приведенное рассуждение показывает, что значения слов, обозначающих действия, нужно рассматривать как структуры, между элементами которых установлены определенные временные соотношения. Но возникает вопрос, можно ли найти и языковые факты, подтверждающие приведенное рассуждение? Оказывается, можно. Здесь мы остановимся только на одной весьма своеобразной конструкции, которая возможна со словами, выражающими целенаправленные действия и которая на первый взгляд скорее опровергает наши рассуждения, чем подтверждает их. Речь идет о таких конструкциях, как:  $s_1$  намеревается просить  $s_2$ , чтобы  $s_2$  сделал  $p$ ;  $s_1$  решил просить  $s_2$ , чтобы он сделал  $p$ . Проблема заключается здесь в следующем: если мы в этих предложениях слово просить просто заменим его семантической структурой (перифразой) в указанном выше смысле, то получается абсурдная интерпретация. Например, интерпретация первого предложения должна быть примерно следующей: "  $s_1$  намеревается желать, чтобы  $s_2$  сделал  $p$ ; он намеревается предполагать, что  $s_2$  сам не собирается сделать  $p$  ...  $s_1$  намеревается принять решение сделать  $p$ ; ..." По-

скольку такая интерпретация является бессмыслицей, то это как будто подтверждает, что информацию об "истории" соответствующего действия было бы неправильно включать в семантическую структуру слов.

Однако это было бы слишком простое решение данной проблемы. Что, собственно, выражают такие конструкции, как намеревается просить, решил уйти, обдумывает, согласиться ли и т.п. На наш взгляд, смысл этих выражений лучше всего можно характеризовать следующим образом: они указывают, что субъект в своей деятельности дошел как раз до того этапа, который называется первым, синтаксически доминирующим глаголом (намереваться, обдумывать, решать). Например, из фразы намеревается просить мы можем вывести всю информацию, которая содержится в семантической структуре слова просить до компонента  $S_1$  INTEND... Таким образом, рассматриваемые фразы нужно интерпретировать как целостные выражения, которые вместе выражают определенную часть семантической структуры соответствующего слова деятельности (просить, соглашаться, требовать и т.п.). Семантическую структуру, соответствующую этим фразам, нельзя конструировать простым введением семантического представления доминируемого слова в структуру доминирующего слова, как это является естественным, например, в случае фразы ударил мяч ногой, где семантические структуры слов нога и мяч просто вставляются в семантическую структуру слова ударить. Но в фразе намеревается просить слово намереваться не служит той "рамой", в которую можно целиком вставить семантическую структуру слова просить. По содержанию здесь доминирующим является глагол просить, а глагол намереваться функционирует как особый "оператор", указывающий, какая часть из структуры слова просить является действительной. На правильность такой трактовки указывает тот факт, что описываемые конструкции могут встречаться только в связи с глаголами, выражающими "субъективные" аспекты деятельности: хотеть, желать, обдумывать, намереваться, решать и т.п., а не с любыми "психологическими" глаголами. Во-вторых, в конструкциях со словами обдумывать, намереваться и др. не возникает указанной проблемы, если их субъект отличается от субъекта соответствующей деятельности. Например, в предложении Мари очень хотелось, чтобы Яан ушел совершенно естественно оперировать семантической структурой глагола уйти как дополнением (объектом) глагола хотеть. Причиной этого

ясна: желание Мари не является "частью" ухода Яана. Следует также указать, что в зависимости от доминирующего глагола мы можем вывести из рассматриваемых фраз разные множества следствий. Например, из фразы намеревается просить можно вывести информацию, что субъект уже решил просить X; предполагает, что посредством рассматриваемой просьбы можно добиться желаемого результата и т.д. Но из фразы думает, просить ли мы можем вывести меньше информации.

Все это показывает, во-первых, что "история" (в описанном выше смысле) некоторой деятельности совершенно объективно относится к значению слова; и во-вторых, что эта информация должна быть упорядочена определенным образом, иначе приведенные выше факты останутся необъяснимыми.

На необходимость упорядочить материал, входящий в значение слова, указывали и раньше, прежде всего в связи с различием утверждения и презуппозиций и установлением определенной иерархии среди последних (Morgan, 1969; Lakoff, 1972).

Но понятие презуппозиции является, на наш взгляд, слишком абстрактным, оно определено слишком формально, чтобы ей можно было придавать какое-нибудь определенное содержание в терминах свойств тех реальных явлений, которые выражаются соответствующим словом, или же в терминах процессов, которые происходят при понимании языка (см. Ёлм, 1977). Установленная нами последовательность имеет, однако, совершенно ясное эмпирическое содержание: это временная последовательность "этапов" деятельности как протекающего во времени процесса. Представленное в таком виде значение слова представляет собой как будто "сценарий", по которому совершается соответствующее действие.

Что касается формального представления такого упорядоченного во времени материала, то мы хотели бы указать, что нужно формализовать такие понятия, выражающие временные соотношения, как "раньше", "позднее", "затем", "одновременно" и др. В лингвистической семантике анализом этих слов (понятий) почти не интересовались. До сих пор этой проблемой занимались логики (Ивин, 1970; Рейор, 1967; Трошкин, 1976 и др.) и, что интересно отметить, исследователи, которые заинтересованы в моделировании целенаправленной деятельности или же в разработке систем, способных понимать тексты, описывающие такую деятельность (Kahn, Gorru, 1977).

## Л и т е р а т у р а

- Звегинцев В.А. Язык и лингвистическая теория. М., "Наука", 1973.
- Звегинцев В.А. Предложение и его отношение к языку и речи. М., "Наука", 1976.
- Ивин А.А. Логика времени. - Неклассическая логика. М., "Наука", 1970.
- Почепцов Г.Г. О некоторых вопросах лингвистики общения. - Структурная и математическая лингвистика, вып. 3. Киев, 1975, 66-73.
- Почепцов Г.Г. Высказывание и трансформация его свойств в неречевых формах общения. - Структурная и математическая лингвистика, вып. 4. Киев, 1976, 54-58.
- Трошкин Е.И. Взаимодействие различных логик времени. - Исследования по логике и методологии науки. М., Изд. МГУ, 1976. 48-57.
- Чумаков Б.И. Логическая концептуализация поведения. - Проблемы моделирования языковой интеракции. Тарту, Изд. ТТУ (в печати).
- Нийм Х., Сильдмяэ И. Решения, действия и язык. - Проблемы моделирования языковой интеракции. Тарту, Изд. ТТУ (в печати).
- Fillmore, Ch. Verbs of Judging: An Exercise in Lexical Semantics. - Studies in Linguistic Semantics. Ed. by Ch. Fillmore and T. Langendoen. New York, 1971.
- Fodor, J.A. Three Reasons for Not Deriving "Kill" from "Cause to Die". - Linguistic Inquiry, 1970, No. 4, 429-438.
- Jackendoff, R. Toward an Explanatory Semantic Representation. - Linguistic Inquiry, 1976, No. 1, 89-150.
- Кас, М.В. Action and Result: Two Aspects of Predication in English. - Syntax and Semantics, vol. 5. New York, 1972, 117-124.
- Kahn, K. and Gorry, G.A. Mechanizing Temporal Knowledge. - Artificial Intelligence, 1977, No.2, 111-134.
- Karttunen, L. Implicative Verbs. - Language, 1971, No.2, 340-358.
- Lakoff, G. Linguistics and Natural Logic. - Semantics of

Natural Language. Ed. by D. Davidson and G. Harman.  
Dordrecht, 1972, 232-296.

McCawley, J.D. Lexical Insertion in a Transformational  
Grammar without Deep Structure. - Papers from the 4th  
Regional Meeting of the Chicago Linguistic Society,  
Chicago, 1968, 71-80.

McCawley, J.D. Kac and Shibatani on the Grammar of Killing.  
- Syntax and Semantics, vol.5. New York, 1972, 125-137.

Prior, A.N. Past, Present and Future. Oxford, 1967.

Schank, R. Conceptual Information Processing. Amsterdam,  
1975.

Schmidt, Ch.F. Understanding Human Action. - Theoretical  
Issues in Natural Language Processing. Cambridge, Mass.  
1975, 215-219.

Scrags, G. A Structure for Actions. Working Paper No. 20.  
Institute of Semantic and Cognitive Studies. Castagnola,  
1975.

Shibatani, M. Three Reasons for not Deriving "Kill" from  
"Cause to Die" in Japanese. - Syntax and Semantics, vol.  
5. New York, 1972, 125-137.

Wierzbicka, A. Why "Kill" does not Mean "Cause to Die"? The  
Semantics of Action Sentences. - Foundations of Language,  
1975, No.4, 491-528.

Wilke, Y. Seven Theses about Artificial Intelligence and  
Natural Language. Working Paper No.17. Institute of Semantic  
and Cognitive Studies. Castagnola, 1975.

Öim, H. On the Semantic Treatment of Predicative Expressions.  
- Generative Grammar in Europe. Ed. by F. Kiefer  
and N. Ruwet. Dordrecht, 1973, 360-383.

Öim, H. Towards a Theory of Linguistic Pragmatics. - Journal  
of Pragmatics, 1977, No.3, 251-268.

Yngve, V. On Achieving Agreement in Linguistics. - Papers  
from the 5th Regional Meeting of the Chicago Linguistic  
Society. Chicago, 1969, 455-462.

Yngve, V. Toward a Human Linguistics. - Papers from the Pa-  
rasession on Functionalism. Chicago, 1975, 540-555.

ON THE ORGANIZATION OF INFORMATION IN THE  
SEMANTIC STRUCTURE OF ACTION VERBS

H. Ginz

S u m m a r y

The paper deals with the problem of semantic analysis of the words designating purposive activities. It is shown that an adequate description of the meanings of these words should include propositions representing (quasi)psychological states and processes which underlie the corresponding action. Two basic "mechanisms" in the model of the agent of an action are more closely dealt with: the mechanism which generates wishes, wants, desires etc., on the one hand, and the mechanism which makes decisions, on the other. Several facts which prove the linguistic reality of these mechanisms are considered.

# SISUKORD - INHALTSVERZEICHNIS - CONTENTS

B o g o d i s t o, V., Information sémasiologique véhiculée par un mot aux sujets parlants et aux étrangers (en russe) .....	5- 13
Résumé en français .....	14
G a s p a r o v, B., P a p e r n o, V., The English Oral Text: A Typological Study .....	15- 31
D u l i ŝ e n k o, A., On the Evolution of the Metalanguage of Present-Day Linguistics (the Prepositive Element LINGVO-) (in Russian) ...	32- 39
Summary in English .....	39
L e h t s a l u, U., L i i v, G., Problems of Rendering the Semantic Structure of the Utterance in Translation .....	40- 49
M u l l a m a a, I., On the Evaluative Colouring of Political Vocabulary (On the Basis of Editorials in "Dagens Nyheter") (in Russian) ...	50- 55
Summary in English .....	56
M u t t, O., Some Notes on Recent and Current Research into English and the Teaching of English (3) .....	57- 64
N a d a r e i s h v i l i, I., O r l o v, Yu., A Method for Complete Fixation of the Text in Linguostatistical Analysis (in Russian) .....	65- 84
Summary in English .....	84
P a a v o, L., Zur Ideologiegebundenheit der neuen Lexik in der DDR und der BRD .....	85- 95
S h a i k e v i t c h, A., Positional Analysis of the Poetic Strophe (in Russian) .....	96-113
Summary in English .....	114
T u l d a v a, J., Eesti keele ühesilbiliste sõnadestruktuuri kvantitatiivne analüüs. - A Quantitative Investigation of the Structure of Estonian Monosyllabics (in Russian) .....	115-134
Summary in English .....	135
U s t a l u, K., Namenkundliche Miscellaneen aus dem 17. Jh. ....	136-139
V i t s o, T.-R., The Non-Uniqueness of Phonological Reconstructions .....	140-146
Õ i l m, H., On the Organization of Information in the Semantic Structure of Action Verbs (in Russian) .....	147-157
Summary in English .....	158



## СО Д Е Р Ж А Н И Е

Б о г о д и с т В. И., Информативность слова для носителей языка и для иностранцев .....	5- 14
Г а с п а р о в В., П а н е р и о В., Английский устный текст: типологическое исследование .... (англ.; рез. рус.)	15- 31
Д у л и ч е н к о А. Д., Эволюция метаязыка современной лингвистики (препозиционный элемент ЛИНГВО-) .....	32- 39
Л е х т е а л у У., Л и й в Г., Проблемы выражения смысловой структуры предложения в переводе .... (англ., рез. рус.)	40- 49
М у л л а м а я И., К вопросу об обеспечении общественно-политической лексики (на материале передовиц "Дагес кхетер") .....	50- 56
М у т т О., Некоторые заметки о современных исследованиях в области английского языка и его преподавания (3) (англ.; рез. рус.) .....	57- 64
Н а д а р е й ш в и л и И. И., О р л о в В. И., Метод полной фиксации текста при лингвостатистическом анализе .....	65- 84
П а л а в о Л., К вопросу об идеологической связанности новой лексики в ГДР и ФРГ .... (нем.; рез. рус.)	85- 95
Ш а й к е в и ч А. Я., Позиционный анализ поэтической строфы .....	96-114
Т у л д а з а Ю., Квантитативное исследование структуры односложного слова в эстонском языке	115-135
У у с т а л у К., Из забытых над ономасной 17 века (нем.; рез. рус.) .....	136-139
В и й т е о Т.-Р., Неоднозначность фонологических реконструкций (англ.; рез. рус.) .....	140-146
И й м Х., Об организации информации в семантической структуре слов, выражающих действия .....	147-158